View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Concordia University Research Repository

EEG-Based Brain-Computer Interfacing via Motor-Imagery: Practical Implementation and Feature Analysis

Mahsa Mirgholami Mashhad

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Electrical Engineering) at

Concordia University

Montréal, Québec, Canada

June 2019

© Mahsa Mirgholami Mashhad, 2019

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Mahsa Mirgholami Mashhad**

Entitled: **EEG-Based Brain-Computer Interfacing via Motor-Imagery: Practical Implementation and Feature Analysis**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Electrical Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair and Examiner
Dr. Hassan Rivaz

_____ External Examiner
Dr. Arash Mohammadi

_____ Examiner
Dr. Habib Benali

_____ Supervisor
Dr. Amir Asif

Approved by _____
Yousef R. Shayan, Chair
Department of Electrical and Computer Engineering

_____ 2019

_____
Amir Asif, Dean
Faculty of Engineering and Computer Science

Abstract

EEG-Based Brain-Computer Interfacing via Motor-Imagery: Practical Implementation and Feature Analysis

Mahsa Mirgholami Mashhad

The human brain is the most intriguing and complex signal processing unit ever known to us. A unique characteristic of our brain is its plasticity property, i.e., the ability of neurons to modify their behavior (structure and functionality) in response to environmental diversity. The plasticity property of brain has motivated design of brain-computer interfaces (BCI) to develop an alternative form of communication channel between brain signals and the external world. The BCI systems have several therapeutic applications of significant importance including but not limited to rehabilitation/assistive systems, rehabilitation robotics, and neuro-prosthesis control. Despite recent advancements in BCIs, such systems are still far from being reliably incorporated within human-machine inference networks. In this regard, the thesis focuses on Motor Imagery (MI)-based BCI systems with the objective of tackling some key challenges observed in existing solutions. The MI is defined as a cognitive process in which a person imagines performing a movement without peripheral (muscle) activation. At one hand, the thesis focuses on feature extraction, which is one of the most crucial steps for the development of an effective BCI system. In this regard, the thesis proposes a subject-specific filtering framework, referred to as the regularized double-band Bayesian (R-B2B) spectral filtering. The proposed R-B2B framework couples three main feature extraction categories, namely filter-bank solutions, regularized techniques, and optimized Bayesian mechanisms to enhance the overall classification accuracy of the BCI. To further evaluate the effects of deploying optimized subject-specific spectra-spatial filters, it is vital to examine and investigate different aspects of data collection and in particular, effects of the stimuli provided to subjects to trigger MI tasks. The second main initiative of the thesis is to propose an element of experimental

design dealing with MI-based BCI systems. In this regard, we have implemented an EEG-based BCI system and constructed a benchmark dataset associated with 10 healthy subjects performing actual movement and MI tasks. To investigate effects of stimulus on the overall achievable performance, four different protocols are designed and implemented via introduction of visual and voice stimuli. Finally, the work investigates effects of adaptive trimming of EEG epochs resulting in an adaptive and subject-specific solution.

Acknowledgments

I send my deepest regards and emotions to my parents and siblings (Maryam, Mehrdad, and Vahid) who have been and are the most precious asset of my life and have been the biggest support for me throughout my life.

I would like to thank my supervisor, Dr. Amir Asif, and express my great appreciation for his guidance, patience, and support during these two years of my Master studies. I wish to express my deep and sincere gratitude and appreciation to my co-supervisor, Dr. Arash Mohammadi, who has made this work possible with his invaluable supports and constructive attitude. Studying under his supervision and working within his team has been a golden opportunity for me. The door to his office was always open whenever I ran into a trouble spot or had a question about my research.

I would like to thank Dr. Habib Benali, who had given me the opportunely to work with him as a research assistant at PERFORM centre. His deep knowledge and astute vision in science has always been an inspiration to me and my work.

I also send my thanks to my dear friends (Afshin, Amirali, Anestesis, Bahareh, Behnoosh, Davoud, Farshad, Farzam, Golnar, Homa, Ida, Kylian, Roya, Saeid, Sarah, Sara) who are a great source of energy and hope to me and have been my second family during my studies and research.

Contents

List of Figures

List of Tables

Chapter 1

Thesis Overview

1.1 Outline

An alarming population ageing is widely expected in near future partially due to recent developments and advancements of biomedical health technologies. According to a recent publication by the United Nations, the number of seniors over the age of sixty is expected to double by 2050; even it is projected that population of seniors will be more than the population of minors/youth at ages 10-24 by 2050. Consequent of this inevitable worldwide population ageing trend is a significant increase in age-related health issues, in particular, neuromuscular disorders. Patients suffering from neuromuscular disorders, typically, have disabilities in performing voluntary movements and controlling their limbs (muscles). Amyotrophic Lateral Sclerosis (ALS), brainstem stroke, and spinal cord injury are potential causes of neuromuscular disorders, which can lead to a state referred to as "locked-in", i.e., the patient's communication with the outer world is completely blocked or limited Kübler, Kotchoubey, Kaiser, Wolpaw, and Birbaumer [2001]. Over recent years, researchers have contemplated on developing new communication technologies for such individuals suffering from neuromuscular diseases J. R. Wolpaw, Birbaumer, McFarland, Pfurtscheller, and Vaughan [2002]. Developed technologies are commonly known as Augmentative and Alternative Communication (AAC) Baxter, Enderby, Evans, and Judge [2012], with the ultimate goal of aiding individuals with communication difficulties to improve their quality of daily life, possibly without requiring caregivers' assistance.

In cases where repair of damages induced by neuromuscular disorders is impossible, the following three options can be pursued to restore (partially or fully) the lost functionalities J. R. Wolpaw et al. [2002]:

(i) The first approach is to increase the capabilities of the remaining sensorimotor systems. For instance, individuals paralyzed by immense brainstem damage can use eye movements to initiate commands, answer questions Kubota et al. [2000], and/or produce speech by hand movement LaCourse and Hludik [1990].

(ii) The second approach is to detour the neural pathways around the damaged area to restore the lost functionality. For example, in spinal cord injuries, control of the paralyzed muscles can be performed by using Electromyographic (EMG) activities from tissues above the level of lesion Kilgore, Peckham, and Keith [1997] and Ferguson, Polando, Kobetic, Triolo, and Marsolais [1999].

(iii) The third option is to directly interface the brain with a computing device to transfer messages and commands to the external world without requiring any peripheral muscular activities.

The latter item (Item (iii)) is commonly referred to as the Brain-Computer Interfacing (BCI) and is the focus of the thesis. Recently, different studies have been conducted for the development of BCI technologies to further improve the quality of life and restore functionality to patients with severe neural injuries and motor disabilities. Application of BCI systems Daly and Wolpaw [2008] for assistive systems can be classified into two major categories:

- In this category, the BCI systems are used to replace the damaged neuromuscular system, i.e., the brain signals are utilized to enable the patient to interact with the outside environment. Consequently, a person can perform a variety of tasks like answering to “yes” or “no” questions, controlling the cursor on a computer or controlling a robotic arm by using electrophysiological (EEG) signals or monitoring the activity of the brain.
- In this category, activity-dependent brain plasticity is induced to restore more normal brain functions, e.g., by focusing on a motor task or by requiring the activation or deactivation of specific brain signals, one could modify brain plasticity.

While the former category is passive, the latter is more active and demands full participation of the patient in the process of interfacing with a computer.

1.2 Objectives

It is shown Edelman, Baxter, and B [2016] that imagination of a movement in peripheral muscles produces a similar effect on the brain rhythms, generated in the sensory-motor cortex as if the real movement is executed. On the grounds of this fact, Motor Imagery (MI) is considered as one of the most standard concepts for the design of BCI systems. The MI is defined as a cognitive process in which a person imagines performing a movement without peripheral (muscle) activation Decety [1996]. The thesis focuses on MI-based BCI systems intending to tackle the following two key challenges observed in existing solutions.

Objective 1. During the practical session of experimenting with MI and EEG-based BCI system, it is observed that a non-negligible portion of users is not able to communicate and work with a BCI system. Generally speaking, while dealing with MI, the BCI developers commonly face two types of challenges:

(i) The first challenge that raises a red flag, is that the subjects of study, typically, comprehend the instructions given on "imagination of the movement" differently. For instance, some subjects imagine repeating the movement during response intervals (epoch), while some others might execute the imagination of the activity only once, and not necessarily consistently within equal time intervals after the stimulus is presented.

(ii) The second challenge is the fact that through a cognitive process, the subject's brain learns to decrease the motor concentration while doing the same task. Hence, the signal's power within the frequency band corresponding to the motory activities tends to descend over time.

In order to tackle the first challenge, one approach is the implementation of subject-specific methods where an adaptive classifier is trained subject by subject. In this case, the processing framework is adaptive to the nature of the datasets collected and is robust by design. For instance, should the subject react to the stimulus right after he/she sees it, the classifier of the dataset collected from

this subject is trained to read the epochs' information immediately after the "marker". The thesis focuses on proposing an alternative and innovative solution to address not only the first but both challenges.

Objective 2. Several studies on different brain imaging modalities have shown that neuro-physiological rhythmic activities (μ and β) recorded over the sensorimotor cortex are modulated by actual movement, and similarly by MI. At the moment of a voluntary movement imagination, a drop in the power of these rhythms is observed, known as event-related Desynchronization (ERD), and once the movement is done, these rhythms emerge again and produce an event-related Synchronization (ERS). Within different approaches developed for MI-based BCI, the common spatial pattern (CSP) is one of the most popular and effective techniques used for classification and feature extraction. The CSP provides spatial filters resulting in more precise detection of the ERD and the ERS waveforms. The CSP technique, however, suffers from sensitivity to noise, over-fitting problem, and is highly dependent on the specification of the utilized frequency bands. To address these concerns, three main research directions were considered in the literature as follows:

(i) At one hand, techniques are developed by using filter banks such as filter bank common spatial pattern (FBCSP) Ang, Chin, Wang, Guan, and Zhang [2012], and separable common spatio-spectral patterns (SCSSP) Aghaei et al. [2016]. The main goal in this category is to form a number of deterministically identified frequency bands.

(ii) The second category Suk and Lee [2013], on the other hand, mainly focuses on optimizing the limits of a single frequency band in contrary to the utilization of several deterministic bands.

(iii) Finally, the third category of research works Lotte and Guan [2011]; Lu, Eng, Guan, Plataniotis, and Venetsanopoulos [2010] focused on regularizing the CSP method to both deal with lack of training data and to have a subject-specific solution.

To take advantage of specific characteristics of each category, hybrid solutions are an attractive alternatives. For example, recently, Reference Park, Lee, and Lee [2018] has combined the first and last categories (Category (i) and (iii)) and proposed a regularized FBCSP framework. Despite potential advantages of hybrid solutions, however, this category is still in its infancy. The thesis further advance this area.

Contributions: Inspired by the stated issues, I have made a number of contributions during my thesis research work as briefly outlined below:

C1. The initiative of this thesis aims at mitigating the challenges mentioned above under Objective 1 by designing an adaptive and detailed search for obtaining the most informative time interval within each epoch. Intuitively speaking, the proposed framework seeks for the best "start-time" and the best "stop-time" within each epoch and readjusts the epochs by trimming the segments before the former and after the latter. In this fashion, the non-informative parts of the epochs are discarded; therefore, no (or less) irrelative time samples are taken into account during the processing module of a BCI. The proposed framework shows an impressive impact on the performance of the MI, EEG-based BCI systems and notably contributes to improving the classification accuracy.

C2. In this thesis, a subject-specific spatio-spectral filtering framework is proposed, referred to as the Regularized double-Band Bayesian (R-B2B) framework. The R-B2B is a hybrid solution that couples all the three aforementioned categories specified under Objective 2 above. More specifically, the original B2B consists of a subject-specific spectral filtering algorithm developed by coupling filterbanks (Category (i)) and optimized techniques (Category (ii)). The B2B optimizes the limits of two bandpass filters, which is computationally far less expensive than the nine filters used in the FBCSP Ang et al. [2012]; Park et al. [2018] and eliminates the need for performing feature selection. The original B2B, however, uses conventional CSP technique for spatial filtering. The R-B2B extends the spatial filtering component of the B2B by incorporating subject-specific spatial filtering, similar in nature to the extension proposed by Reference Park et al. [2018] over the conventional FBCSP Ang et al. [2012].

C3. Another initiative of this thesis is to propose an element of experimental design dealing with MI-based BCI systems, in order to provide a benchmark dataset to further advance EEG signal processing techniques. In other words, to further examine the effects of deploying optimized subject-specific spectra-spatial filters, it is vital to examine and investigate different aspects of data collection and in particular, effects of the stimuli provided to a subject to trigger actual motion or MI tasks. In this regard, we have started collecting EEG signals based on

a portable and reliable wireless biosignal acquisition system, g.Nautilus from g.tech Medical Engineering. Several experiments are performed, and a dataset is formed from EEG signals associated with 10 healthy subjects during actual movement and MI tasks. To investigate the effects of stimulus on the overall achievable performance, four different protocols are designed and implemented via introduction of visual and voice stimuli.

1.3 Organization of the Thesis

To provide the relevant context, the rest of the thesis is organized as follows:

- Chapter 1 provides an overview and a summary of important contributions made in the thesis.
- Chapter 2 provides a comprehensive background and literature review on BCI systems and their applications, as well as the constituent modules and the corresponding feature extraction methods.
- Chapter 3 presents experimental data collection protocol and collecting Electroencephalography (EEG) signals from 10 healthy volunteer subjects performing Actual Movement (AM) and Motor Imagery (MI) tasks among the different suggested protocols.
- Chapter 4 introduces three practical solutions proposed to increase the classification accuracy of BCI systems. In brief, the proposed solutions are as follows: (i) Improving the accuracy of EEG-based BCIs based on Tuning Algorithm Parameters, stimulation and results for cross-Validation algorithm; (ii) Improving the Performance of MI and EEG-based BCIs via an Adaptive Epoch Trimming Mechanism, and (iii) Adaptive subjected-specific Bayesian spectral filtering for signal trial EEG Classification.
- Chapter 5 concludes the thesis and provides some directions for future research.

Chapter 2

Background and Literature Review on Brain Computer Interfacing

The human brain is the most intriguing and complex signal processing unit ever know to humans. A unique characteristic of our brain is its plasticity property, i.e., the ability of neurons to modify their behavior (structure and functionality) in response to environmental diversity. This exclusive property has recently been utilized by researchers, to recover the ability of moving legs in patients paralyzed with spinal cord injuries after training them with an exoskeleton linked to their brain. The plasticity property of brain has also motivated the design of brain-computer interfaces (BCI) to develop an alternative form of communication between human brain signals and the external interfacing world. BCIs have several therapeutic applications of significant importance to help partially or fully-disabled patients interact with outside world, including but not limited to rehabilitation/assistive systems, rehabilitation robotics, and neuro-prosthesis control. Despite recent advancements in BCIs, such systems are still far from being reliably incorporated with human-machine inference networks. The remainder of the chapter is organized as follows: In Section 2.1 fundamentals of Brain Computer Interfaces (BCIs) are reviewed. A literature review on main feature extraction approaches used in BCI systems is presented in Section 2.2.

2.1 Fundamentals of Brain Computer Interfaces

Let's start with answering this question: What is a BCI? Brain-computer interfacing is an emerging communication system, which acts as a communication language between human brain signals and computers Kübler et al. [2001]. In other words, a BCI can be considered as a system in which the input is brain activities, and the output is a set of control/actuation commands Vidal [1973]. The BCIs have found roots in a wide variety of applications including but not limited to the following categories:

(1) Medical Applications;
(2) Nuroergonomics and Smart Environment;
(3) Neuromarketing and Advertisement;
(4) Educational and Self-regulation;
(5) Games and Entertainment, *and*;
(6) Security and Authentication.

Generally speaking, BCI systems have potential applications in medical domain for *Prevention*, *Detection*, *Diagnosis*, and *Rehabilitation* purposes. Specifically, mobility rehabilitation, which is a form of physical rehabilitation aiming to restore the lost function of patients, who have mobility issues, or assist them in doing daily activities independently Van Erp, Lotte, and Tangermann [2012]. There have been numerous research studies on BCI systems to improve the quality of life for the people with neural injuries and motor disabilities. For rehabilitation purposes Daly and Wolpaw [2008], at one hand, BCI can be used for enabling patients to interact with their environment, e.g., performing tasks such as answering to Yes/No questions to control a cursor on computer's screen or control a robotic arm. On the other hand, the focus of BCI systems is on restoring motory task by inducing activity-dependent brain plasticity to restore near-normal brain function. Next, the main components of a BCI system are briefly introduced.

Among different type of brain signals deployed to control the movement, three Yuan and He [2014] are of more interest for BCIs, i.e., electrophysiological signals recorded over the scalp; over the cortical surface (electrocorticography), and; within the brain (single-neuron action potentials (single units) and local field potentials (LFPs)). The common aspect of these methods is that they measure the extracellular potentials generated by the neurons in cortical layers, however, they differ

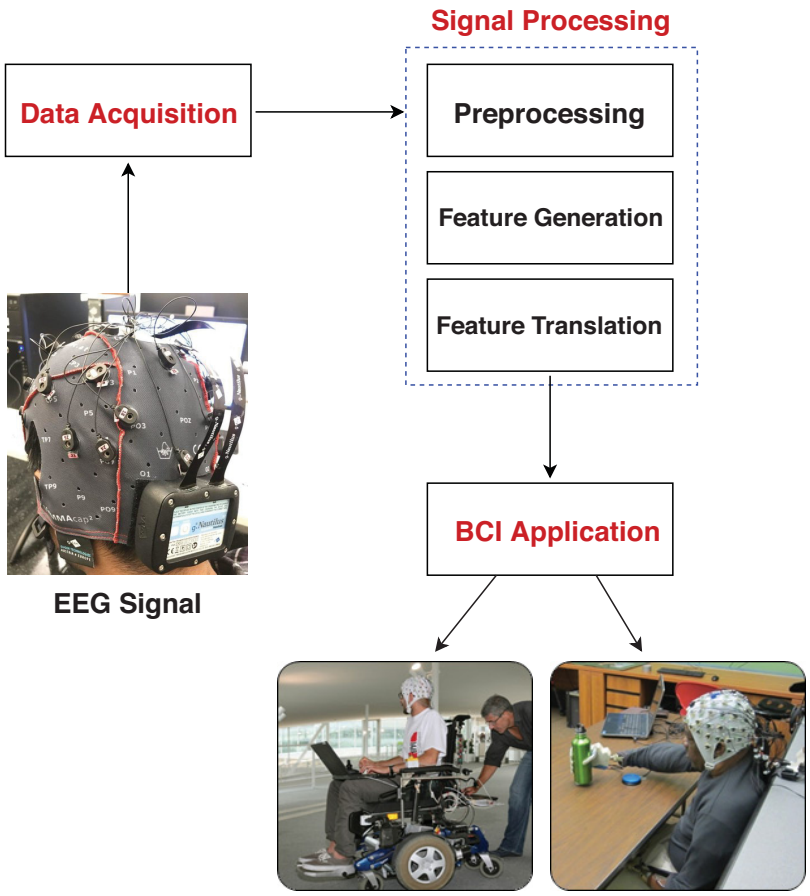

Figure 2.1: Essential blocks of a BCI system.

in the field distance and spatial resolution. Some research works have uncovered how the kinematic parameters of the movement are encoded in the neuronal electrical activities. Thanks to these findings, researchers have been able to develop real-time closed-loop BCI systems. Initially, these systems were tested on nonhuman primates, and then electrode arrays were implanted in the severely disabled patients for multidimensional control of a computer cursor or a robotic arm. Intracortical recordings (mostly single units) facilitate achieving a high level of degree-of-freedom (DOF) on the BCI systems, however, the long-term stability of the electrodes is still a matter of concern, which limits their clinical applications.

A BCI system, as shown in Fig. 2.1, typically consists of two main components, i.e., *Data Acquisition Module*, and *Signal Processing Module*. The former is about recording the brain activities from the brain, while the latter is about pre-processing the recorded signals in terms of extracting

informative features, Fig. 2.2. Next, I briefly review these two categories.

Signal Acquisition Module: As the first vital block of any BCI system, a modality is required to record brain activities; thus, different modalities have been introduced, and their effectiveness has been investigated. Acquisition of brain waves by a BCI system is required to infer its neural activities and can be performed in either of the following two different ways Moore [2003]: *(i) Invasively* Nijholt et al. [2008] from the electrodes implanted surgically inside the patient's brain. Implanted electrodes aim at sending/collecting signals directly from the brain, therefore, provide high quality brain responses Chen et al. [2009] and Postelnicu, Talaba, and Toma [2010], and; *(ii) Noninvasively* Zander and Kothe [2011] by recording brain signals from the external surface of the scalp without any surgical implantation. Such techniques are the most extensively researched modalities considering their accessibility in organizing studies and their minimal risk.

With regards to non-invasive acquisition of brain activities, *Electroencephalography (EEG)* Moghimi, Kushki, Marie Guerguerian, and Chau [2013] is considered as the most popular modality due to offering several advantages including higher temporal resolution, flexibility in design, possibility to change electrodes' placements, and accessible cost in comparison to other neuroimaging techniques. In recent years, several clinical studies, ranging from neuroscience, cognitive science, cognitive psychology, neurolinguistics, and psychophysiological, have been performed on EEG-based BCI systems approving its potential effectiveness.

To develop an EEG-based BCI system Yuan and He [2014], one option is to modulate neurophysiological activities over the sensory motor cortex by performing an actual movement or *Motor Imagery (MI)* movements. The latter (i.e., MI) is defined as a cognitive process in which a person merely imagines a limb movement without any actual or peripheral (muscle) activation Decety [1996]. It is shown in Jeannerod [1995] that mental imagination of movement involves similar brain regions as an actual movement. Motor imagery tasks modulate Sensorimotor Rhythms (SMRs), which can be used for classification of different actions. However, since the EEG recording is performed over the extracellular field potentials over the scalp, its spatial resolution is limited (in centimetres level) and the recordings are severely prone to be interfered by environmental and physiological artifacts, e.g., electromyographic signals from cranial muscles or electrooculographic

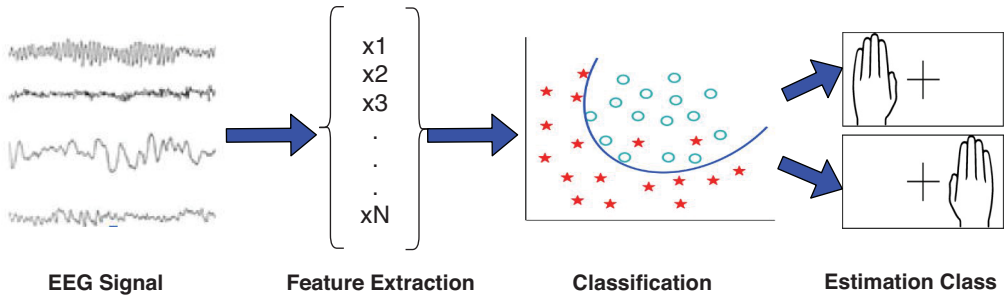

Figure 2.2: A classical EEG signal processing pipeline for BCI.

activity. This calls for the development of spatial filtering and spatial processing techniques to improve the overall calcification accuracy of the system, which in part is the focus of this thesis. In the following sub-sections, I review the basics of EEG signals followed by the description of EEG processing pipeline.

2.1.1 EEG Signals

Generally speaking, EEG signals are collected by placing electrodes on the scalp measuring the voltage changes within the neural movement of the brain. The values of EEG recordings are typically small in the range of micro voltage resulting in the collected EEG signals to consist of different artifact sources introduced for example by blinking and muscle contraction. Signals associated with brain activities fall between 1-40 Hz and frequency contents below or above this range are considered as artifacts. Rhythmic activities observed in EEG signals can be divided into different frequency bands associated with varying states of the brain. A brief explanation of the main frequency bands of the EEG waves Abo-Zahhad et al. [2015] are provided below:

- *Delta Waves (1-4 Hz)*: Which is the high amplitude brain waves and considers as a slow sleep stage. Moreover, these signals are widely studied in cognitive processing captured from the frontal part of the scalp.

- *Theta Waves (4-8 Hz)*: Normally considered as an "idling" waveforms and mostly found at the location of the scalp that is not related to the task at hand. Furthermore, they have a significant role in attentional processing and work memory.

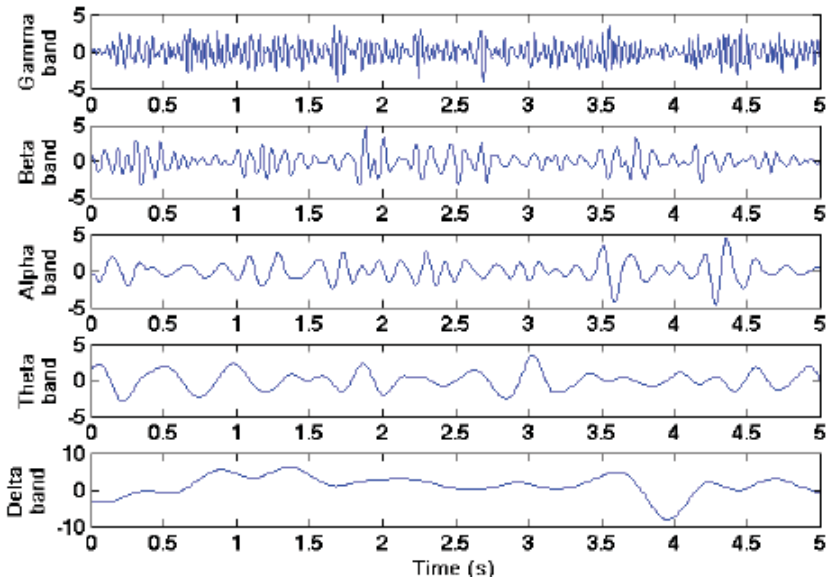

Figure 2.3: The five frequency bands of EEG signal. Abo-Zahhad et al. [2015]

- *Alpha Waves (8-13 Hz)*: It is also known as μ rhythm associated with a state of relaxing and resting. During the eyes-closed condition, alpha waves are well known in the posterior region of the head.

- *Beta Waves (14-30 Hz)*: This category is of low amplitude and it is generated during activeness, anxious thinking, problem-solving, and deep concentration at the frontal and central regions.

- *Gamma Waves (Above 30 Hz)*: Which is found in the somatosensory cortex when two different senses are combined. Moreover, it is involved in attention, working memory, and long-term memory processes.

Preparation for movement results in a decrease in power of Alpha and Beta rhythms, which is named *Event Related Desynchronization (ERD)*. Moreover, after the movement, Alpha and Beta rhythms will be increased called *Event-Related Synchronization (ERS)* as shown in Figs. 2.5 and 2.4.

2.1.2 EEG Signal Processing

EEG Signal Processing: Collected EEG signals are analyzed through the BCI processing pipeline consisting of the following main components:

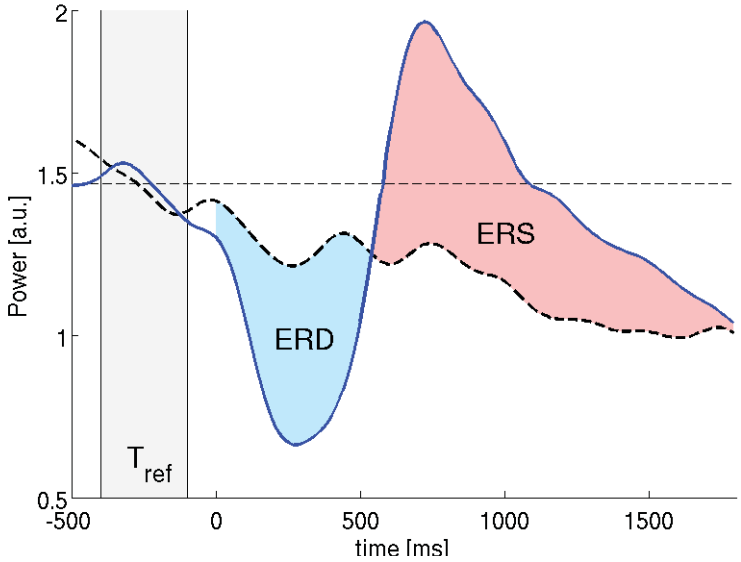

Figure 2.4: ERS and ERD at sensory motor area during tasks. Lemm et al. [2009]

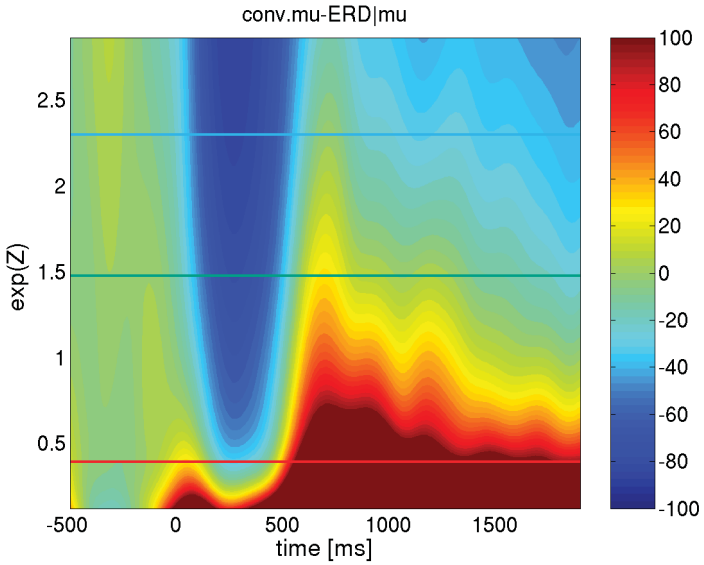

Figure 2.5: ERS and ERD at sensory motor area during tasks. Lemm et al. [2009]

(i) *Preprocessing*: Recorded signals from an EEG channel contains different unwanted sources such as artifacts, brain signals, and noise. In this step, therefore, the collected signals are first filtered to extract specific frequency contents of the signals.

(ii) *Feature Extraction*: An important step aiming to extract specific and relevant information from the EEG recording. Among different processing techniques, *Common spatial patterns*

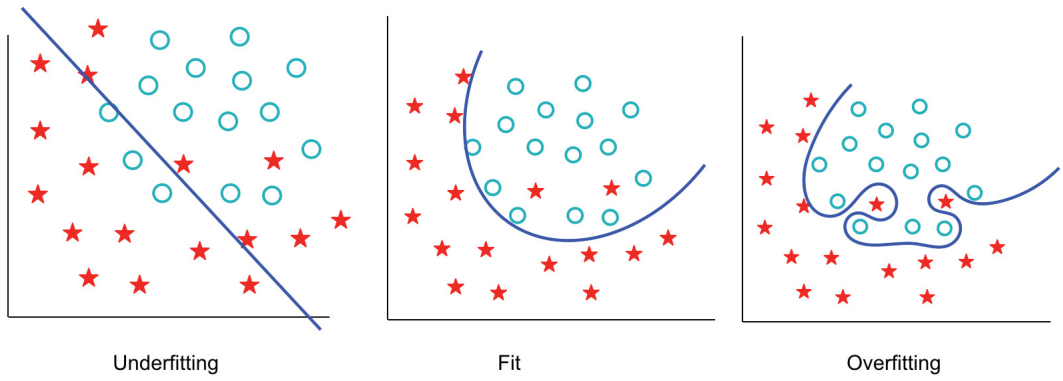

Figure 2.6: The bias-variance trade-off problem, i.e., problem of underfitting and overfitting.

(CSP) is considered as one of the most popular and effective tools for extracting discriminative features to be used for development of MI-based BCI systems. The CSP, as a mathematical algorithm, introduces spatial filters for multichannel EEG signals that aim to find a way to maximize the difference in variance between to signal classes. In other words, the CSP algorithm uses a linear transformation to maximize the variance in one class while the variance in other class is minimized Mason and Birch [2003]. Different variants of CSP feature extraction method will be discussed later in the Section 2.2.

(iii) *Signal classification*: This step involved the process of translating the extracted features into class labels. In other words, this module aims to learn about the relation between features and a target of interest. Classification algorithm starts with choosing a suitable model compatible with the feature vectors. After choosing the optimal model of a classifier, the parameters of the chosen model are determined and the training data are used to improve the classification model's feature prediction accuracy. Once the training is done, the trained model will be validated, and finally, the evaluated model is applied to the test set estimating how accurately it will perform in practice.

Classifiers have been mainly concerned with facing the bias-variance trade-off problem Raudys and Jain [1991], as shown in Fig. 2.6. Low-performance classifiers may occur as a result of not proper fitting of the classification model in terms of simplicity or complicity of the model to describe the target. Basically, *Underfitting* occurs when the model or the algorithm does not fit the

data well enough, i.e., if the model shows low variance but high bias, especially if the model is excessively simple. On the other hand, *Overfitting* appears when a statistical model captures the noise of data. Intuitively speaking, overfitting occurs when the model or the algorithm fits data too well. Particularly, this is often the result of an excessively complicated model and shows low bias but high variance.

In terms of dealing with classification of two feature sets, *Linear and Quadratic Discriminant Analysis (LDA and QDA)* together with *Support Vector Machines (SVM)* are selected as the classification models. Below, a brief outline of each these models is provided:

- **Linear and Quadratic Discriminant Analysis:** This category is a conventional classification model Scherer, Muller, Neuper, Graimann, and Pfurtscheller [2004], which is designed for two class problems but can be modified to be applied for multiple class problems Garrett, Peterson, Anderson, and Thaut [2003]. A linear discrimination function assumption is based on multivariate Gaussian distribution and covariance of the predicted variable are common across all levels of the responses (labels). The LDA has very low computational requirements making it the most commonly used classifier for BCI system. However, we can state that a QDA is a generalization of the linear model, which is assumed that the observation within each class is drawn with a normal distribution. Unlike the LDA classifier, in the QDA, the covariance matrix of each class is identical that involved the quadratic terms.

- **Support Vector Machines:** A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, it finds out an optimal hyper-plan space or line in terms of separating two classes. An SVM selects the optimal line that maximizes the distance between the nearest training samples and the hyperplane Burges [1998] and has been successfully used in several BCI applications Subasi and Gursoy [2010].

2.2 Literature Review on Feature Extraction

As discussed previously, different mental activities result in different patterns of brain signals and BCI is seen as a pattern recognition system by classifying each pattern into a class according to the extracted features. A feature is a measurable property or characteristic of that phenomenon,

which describes some relevant properties of the signals and, if selected suitably, will contribute to rather an accurate classification of the entire dataset. In other words, feature extraction can be considered as a crucial step in the design of BCI systems.

It is difficult to classify recorded EEG signal accurately Sun and Zhang [2006], mainly because of high variability of EEG recordings due to subject attention, subject fatigue, disease progression, electrode impedances, amplifier noise, and environmental noise Vaughan et al. [2003]. The main challenge here is to define and extract discriminative features. Next, different CSP-based spatial filter design approaches are reviewed including *Regularized CSP*; *Filter Bank Common Spatial Pattern (FBCSP)*; *Separable Common Spatio Spectral Patterns (SCSSP)*; *Bayesian spatio-spectral filter optimization (BSSFO)*, and; *Double-band Bayesian Spatio-spectral Filter Optimization*.

2.2.1 CSP-based Methodologies

The CSP methodology is a spatial filtering and dimension reduction design, which is one of the most known and well-regarded techniques for EEG signal processing. The main idea of the CSP approach is to utilize a linear transform for the purpose of projecting the multi-channel EEG data into a new space with the same number of channels in a way that magnifies the underlying features related to the particular task at hand. The CSP algorithm uses a linear transformation matrix to maximize the variance in one class, while the variance in the other class is minimized Xie, Yu, Lu, Gu, and Li [2017]. Moreover, the CSP technique uses the labels of the data and handles the problem in a supervised fashion, therefore, provides superior results in comparison to other analytical techniques such as Principal Component Analysis (PCA) Yu, Chum, and Sim [2014], or Independent Component Analysis (ICA).

Throughout this section, the following notations are used: Considering supervised learning from EEG signals based on the available set of EEG epochs (trials) denoted by

$$\boldsymbol{X}_i \in \mathbb{R}^{N_{\text{ch}} \times N_{\text{t}}}, \qquad \text{for } (1 \leq i \leq N_{\text{Trial}}), \tag{1}$$

where

- N_{Trial}, is the total number of epochs used for processing;

- N_{ch} denotes the number of EEG channels (electrodes), and;
- N_{t} is the number of time samples collected from each electrode in each trial.

Trial $\boldsymbol{X}_i$ is obtained from bandpass filtering of an EEG signal. The training dataset is denoted by $\{(\boldsymbol{X}_i, L_i)\}$, for $(1 \leq i \leq N_{\text{Trial}})$, where L_i represents the label corresponding to the i'th trial. To apply the spatial filters based on the CSP methodology, initially, the spatial normalized covariance matrix for trial $\boldsymbol{X}_i$ is computed as follows

$$\boldsymbol{\Sigma}_i = \frac{1}{N_{\text{t}} - 1} \big(\boldsymbol{X}_i - \boldsymbol{\mu}_i\big)\big(\boldsymbol{X}_i - \boldsymbol{\mu}_i\big)^T. \tag{2}$$

When all classes have zero mean, the normalized spatial covariance matrix is given by

$$\boldsymbol{\Sigma}_i = \frac{\boldsymbol{X}_i \boldsymbol{X}_i^T}{\text{Tr}(\boldsymbol{X}_i \boldsymbol{X}_i^T)}. \tag{3}$$

By keeping in mind that the CSP methodology is designed for discrimination of two classes of data, $\boldsymbol{S}^{(0)}$ and $\boldsymbol{S}^{(1)}$ are used as two sets containing the training trials from each class, $c \in \{0, 1\}$ and the average of spatial covariance matrices of different trials belonging to each set ($\bar{\boldsymbol{\Sigma}}^{(0)}$ and $\bar{\boldsymbol{\Sigma}}^{(1)}$) are computed, then the composite spatial covariance matrix denoted by $\boldsymbol{\Sigma}^{(c)}$ is computed as

$$\boldsymbol{\Sigma}^{(c)} = \bar{\boldsymbol{\Sigma}}^{(0)} + \bar{\boldsymbol{\Sigma}}^{(1)}. \tag{4}$$

The next step is to perform an eigenvalue decomposition on the composite covariance matrix as follows

$$\boldsymbol{\Sigma}^{(c)} = \boldsymbol{V}^{(c)} \boldsymbol{\lambda}^{(c)} [\boldsymbol{V}^{(c)}]^T, \tag{5}$$

where $\boldsymbol{V}^{(c)}$ is the matrix containing eigenvectors associated with the composite covariance and $\boldsymbol{\lambda}^{(c)}$ is the diagonal matrix of its corresponding eigenvalues. Based on Ang et al. [2012] and using Matlab notation, the spatial filter is computed as

$$\boldsymbol{W} = \text{eig}(\bar{\boldsymbol{\Sigma}}^{(0)} + \bar{\boldsymbol{\Sigma}}^{(0)}, \bar{\boldsymbol{\Sigma}}^{(1)}), \tag{6}$$

Algorithm 1 Common Spatial Patterns Step by Step

Input: Available set of EEG epochs (trials) denoted by $\{\boldsymbol{X}\}_{i=1}^{N_{\text{Trial}}}$, size $N_{\text{ch}} \times N_{\text{t}} \times N_{\text{Trial}}$ and corresponding labels $\{\Omega\}_{i=1}^{N_{\text{Trial}}}$
Output: Feature Vectors, ready to be classified.

1: Covariance Matrices: Computing the sample covariance matrix corresponding to each trial $\boldsymbol{X_i}$ for two sets of training set $\boldsymbol{S}^{(0)}$ and $\boldsymbol{S}^{(1)}$

2: Whitening Matrix: Perform spatial filtering by means of Eq. (7) and Select the first and last m rows of the spatially filtered signals.

3: Finalizing the feature vectors: by using Eq. (8)

where the eig$(.)$ function performs eigenvalue decomposition on the two input matrices. The spatial filter $\boldsymbol{W}$ is then used to form the decomposition of each trial $\boldsymbol{X}_i$ for $(1 \leq i \leq N_{\text{Trial}})$ as

$$\boldsymbol{\mathcal{W}}_i = \big[\boldsymbol{W}\big]^T \boldsymbol{X}_i. \tag{7}$$

For each direction of feature, $\boldsymbol{\mathcal{W}}_i$ is computed and for discrimination analysis, only the first and last m rows of the matrix are used for the construction of the classifier. In other words, matrix $\boldsymbol{\mathcal{W}}_{i,m}$ is constructed from the first and last m rows of the matrix $\boldsymbol{\mathcal{W}}_i$ which represents rows of $\boldsymbol{\mathcal{W}}_i$ associated with the largest eigenvalues that maximizes the difference of variance between the two classes. The following feature vector is then used within the classification step

$$\boldsymbol{f}_{i,m} = \log\left(\frac{\text{var}(\boldsymbol{\mathcal{W}}_{i,m})}{\sum_{k=1} \text{var}(\boldsymbol{\mathcal{W}}_{i,m})}\right). \tag{8}$$

Note that, the log-transformation in Eq. (8) is included to magnify the distance of the features.

2.2.2 Regularized Common Spatial Pattern (RCSP)

Despite providing several advantages, the CSP approaches continue to suffer from sensitivity to noise and adverse over-fitting effect with small training samples Reuderink and Poel [2008] Grosse-Wentrup, Liefhold, Gramann, and Buss [2009]. Since CSP framework is based on sample-based covariance-matrix estimation and the covariance matrix will be estimated from the training samples, the small training sample becomes a practical EEG classification problem leading to poorly estimated covariance matrix and classification accuracy Lu et al. [2010].To address this drawback,

different studies Lotte and Guan [2011] have been conducted to deal with the small training set problem by regularizing the CSP methodology based on covariance estimation level.

In the regularized CSP, first the covariance matrix for each trial $\boldsymbol{X}_i$, $(1 \leq i \leq N_{\text{Trial}})$, is used for discrimination of two classes and is computed as $\boldsymbol{\Sigma}_i = (\boldsymbol{X}_i \boldsymbol{X}_i^T)/(\text{Tr}(\boldsymbol{X}_i \boldsymbol{X}_i^T))$. The generic average spatial covariance matrix for each class is then calculated from

$$\boldsymbol{\Sigma}^{(c)} = \frac{1}{N_{\text{Trial}}^{(c)}} \sum_{i=1}^{N_{\text{Trial}}^{(c)}} \boldsymbol{\Sigma}_i^{(c)}, \qquad c \in \{0, 1\}, \tag{9}$$

where $N_{\text{Trial}}^{(c)}$ is the number of epochs/trials belonging to class c, $(c \in \{0, 1\})$. We also compute a subject-specific average covariance matrix denoted by $\boldsymbol{\Sigma}^{(c,j)}$ for subject j, $(1 \leq j \leq N_{\text{Sub}})$, with N_{Sub} denoting the total number of available subjects. Note that, $\boldsymbol{\Sigma}^{(c,j)}$ is computed from Eq. (9), where the summation is performed over the number of epochs available for Subject j. The regularized subject-specific covariance matrix is then computed as follows

$$\hat{\boldsymbol{\Sigma}}^{(c,j)}(\beta, \gamma) = (1 - \gamma)\hat{\boldsymbol{\Sigma}}^{(c,j)}(\beta) + \gamma / N_{\text{ch}} \, \text{tr}\big[\hat{\boldsymbol{\Sigma}}^{(c,j)}(\beta)\big] \boldsymbol{I}. \tag{10}$$

In Eq. (10), $\boldsymbol{I}$ is an $(N_{\text{ch}} \times N_{\text{ch}})$ identity matrix and $\hat{\boldsymbol{\Sigma}}^{(c,j)}(\beta)$ is defined as

$$\hat{\boldsymbol{\Sigma}}^{(c,j)}(\beta) = \frac{(1 - \beta)\boldsymbol{\Sigma}^{(c,j)} + \beta \boldsymbol{\Sigma}^{(c)}}{(1 - \beta) N_{\text{Trial}}^{(c,j)} + \beta N_{\text{Trial}}{}^{(c)}}, \qquad c \in \{0, 1\}. \tag{11}$$

Determining the regularization parameters is one important task in RCSP framework Lu, Platanios, and Venetsanopoulos [2009] and Friedman [1989]. In other words, selecting a good value for two regularization parameters lead to have a higher classification performance. In this regard, regularization parameter combinations (where β and γ are regularization parameters between interval $[01]$), were tested and the best result for each case was reported using cross-validation method. For Subject j, the composite spatial covariance matrix is computed as

$$\bar{\boldsymbol{\Sigma}}^{(j)}(\beta, \gamma) = \hat{\boldsymbol{\Sigma}}^{(0,j)}(\beta, \gamma) + \hat{\boldsymbol{\Sigma}}^{(1,j)}(\beta, \gamma). \tag{12}$$

The next step is to perform an eigenvalue decomposition on the regularized composite spatial covariance matrix for subject j, ($1 \leq j \leq N_{\text{Sub}}$), Following by

$$\bar{\boldsymbol{\Sigma}}^{(j)}(\beta, \gamma) = \boldsymbol{V}^{(j)} \boldsymbol{\lambda}^{(j)} [\boldsymbol{V}^{(j)}]^T, \tag{13}$$

where $\boldsymbol{V}^{(j)}$ is the matrix containing eigenvectors associated with the composite covariance and $\lambda^{(j)}$ is the diagonal matrix of its corresponding eigenvalues. Based on Ang et al. [2012] and using Matlab notation, the spatial filter is computed as

$$\boldsymbol{W}^{(j)} = \text{eig}(\bar{\boldsymbol{\Sigma}}^{(j)}(\beta, \gamma), \hat{\boldsymbol{\Sigma}}^{(1,j)}(\beta, \gamma)). \tag{14}$$

The spatial filter $\boldsymbol{W}_i^{(j)}$ is then used to form the decomposition of each trial $\boldsymbol{X}_i$,($1 \leq i \leq N_{\text{Trial}}$). $\boldsymbol{\mathcal{W}}_i$ is constructed from the first and last l rows of matrix $\boldsymbol{W}^{(j)}$ which is associated with the largest eigenvalues that maximizes the difference of variance between the two classes is computed as

$$\boldsymbol{\mathcal{W}}_{i,l}^{(j)} = [\boldsymbol{W}_l^{(j)}]^T \boldsymbol{Z}_{i,l}^{(k)}. \tag{15}$$

Finally, variance operator, donated by $\boldsymbol{f}_i$ is included to magnify the distance of the features, i.e.,

$$\boldsymbol{f}_{i,l} = \log \left(\frac{\text{var}(\boldsymbol{\mathcal{W}}_{i,l})}{\sum_i \text{var}(\boldsymbol{\mathcal{W}}_{i,l})} \right). \tag{16}$$

Similar to the obtained feature vector in Eq. (8), $\boldsymbol{f}_{i,l}$ will be used for the classification step.

2.2.3 Filter Bank Common Spatial Patterns (FBCSP)

The CSP algorithm is effective due to building up optimal spatial filters discriminating two different class of EEG measurements. However, several parameters, such as frequency for band-pass filtering and time interval of the EEG measurements, have to be specified. In particular, feature extraction of EEG signal using CSP hugely depends on the selection of the frequency bands. Since the frequency bands are subject-specific, it is difficult to determine the optimal filter bands Kumar, Sharma, and Tsunoda [2017]. Several studies; Blankertz, Tomioka, Lemm, Kawanabe, and

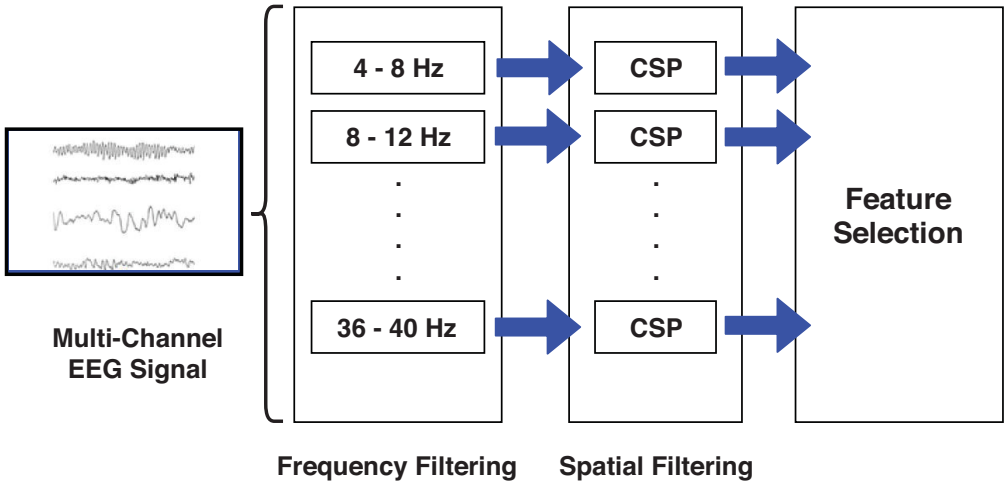

Figure 2.7: Principle of Filter Bank Common Spatial Patterns (FBCSP). 1- Dividing Raw EEG signals into the smaller frequency band using filter bank; 2- Optimizing CSP spatial filter for each band; 3- Selecting the most relevant filters (both spatial and spectral) using feature selection.

Muller [2008], and Ang et al. [2012] have suggested that optimizing the filter band could improve the accuracy of EEG signal classification. The *Filter Bank Common Spatial Patterns (FBCSP)* algorithm (shown in Fig. 2.7) is performed to enhance the performance of the CSP algorithm. This method selects some bandpass filter and comprises of four different stages of signal processing on the EEG data, i.e., multiple bandpass filtering; spatial filtering using CSP; feature selection of the CSP features, and; classification.

In particular, in the first stage, EEG data (4-40 Frequency range) is decomposed into multiple smaller frequency bands. In the second stage, spatial filtering by using the CSP algorithm applies to each component for feature extraction. In the last step, the FBCSP deploys a feature selection method to select a subset of features, which best represent the class labels in the training phase, and finally feeds them into the classier for training purposes.

2.2.4 Separable Common Spatio-spectral Patterns (SCSSP)

Despite the CSP power in extracting spatial features, CSP completely ignores the spectral characteristics of the EEG signals. *Separable Common Spatio-spectral Patterns (SCSSP)* technique Aghaei et al. [2016] is a novel algorithm, which discriminate spatio-spectral features across all the frequency band. In other words, the SCSSP processes the EEG activities simultaneously in

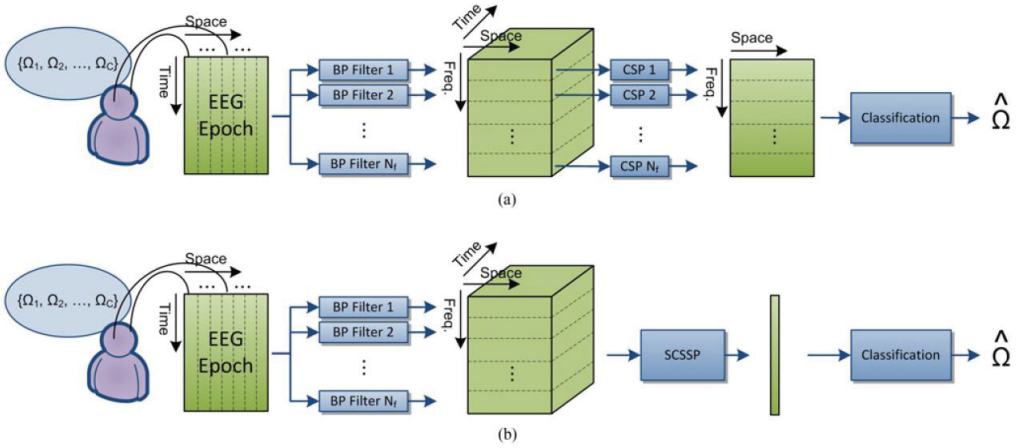

Figure 2.8: System model for spatio-spectral feature extraction schemes in (a) Filter-bank common spatial pattern (FBCSP), and (b) Separable common spatio-spectral pattern (SCSSP) methods Aghaei et al. [2016].

both spatial and spectral domains. The SCSSP method is an extension to the FBCSP by developing a bilinear feature extractor. The SCSSP has significantly less computational cost for training in comparison to the FBCSP method since it requires training of only two CSP-type modules. Moreover, the features are extracted based on the joint analysis of both spatial and spectral characteristics of the signal. In particular, the correlation between different spectral bands is exploited for the feature extraction step.

In the SCSSP method, shown in Fig. 2.8, the EEG raw signal consist of $N_{\text{Trial}} \times N_{\text{t}}$ from N_{ch} channels. First of all, different rhythmic activities from each EEG channel are extracted by applying a set of N_{f} bandpass filters. Each rhythm has a length of N_{Trial} matrices of size $N_{\text{f}} \times N_{\text{ch}}$. Moreover, each of these matrices represents spatio-spectral EEG patterns at a certain time instant. The SCSSP algorithm applies to these spatio-spectral EEG patterns to extract features for classification by using a heteroscedastic matrix-variate Gaussian model Harrar and Gupta [2008]. The SCSSP method takes into account the joint characteristics of the spatial and spectral features. In other words, this framework processes the data in both spectral and spatial domains and hence can sort the extracted features across both domains and reduces its computational cost.

2.2.5 The Bayesian Spatio-spectral Filter Optimization Framework (BSSFO)

As discussed previously, by taking advantage of a filter bank, both FBCSP and SCSSP frameworks decompose the EEG recordings into several spectral components and extract frequency-specific features, which has resulted in high classification accuracy. However, performing feature extraction from a non-optimal frequency band (which happens during the optimization iterations) could lead to miss potentially informative features. Reference Suk and Lee [2013] proposed expanding optimized spectral filters rather than using filter banks. This research deployed a novel framework of Spatio-spectral filter optimization for discriminative feature extraction. This framework is divided into three steps, i.e., *Spectral Filtering*, *Spatial Filtering*, and *Feature Extraction* as are described below

Spectral Filtering: This step is modeled as the convolution of the EEG trial $\boldsymbol{X}_i$ with the impulse response of a spectral bandpass filter h given by

$$\boldsymbol{\mathcal{Z}}_i = h \circledast \boldsymbol{X}_i. \tag{17}$$

The cutoff frequencies of a bandpass filter (h) are defined as $\boldsymbol{\mathcal{B}} \triangleq [b_s, b_e]$, where b_s and b_e are, respectively, the start and the end frequency of the band range. The chance that bandpass-filtered signals can be correctly classified between two classes is defined by the probability of a frequency band and the uncertainty in the cutoff frequencies of the spectral filters is modeled with a prior probability denoted by $p(\boldsymbol{\mathcal{B}})$ over random variable $\boldsymbol{\mathcal{B}}$. In particular, the prior distribution $p(\boldsymbol{\mathcal{B}})$ describes relative probabilities of different frequency bands for the set of single-trial EEG $\boldsymbol{X}$. The posterior density is then computed as follows

$$p(\boldsymbol{\mathcal{B}}|\boldsymbol{X}, \Omega_i) = \frac{p(\boldsymbol{X}, \Omega_i|\boldsymbol{\mathcal{B}})p(\boldsymbol{\mathcal{B}})}{p(\boldsymbol{X}, \Omega_i)}, \tag{18}$$

where $p(\boldsymbol{X}, \Omega_i|\boldsymbol{\mathcal{B}})$ is called the likelihood function. If the frequency band (i.e., of hypothesis $\boldsymbol{\mathcal{B}}$) were true, the likelihood function indicates the probability that the single-trial EEG $\boldsymbol{X}$ in conjunction with the class labels Ω would have been available to support it. The probability of frequency

band $\boldsymbol{\mathcal{B}}$ being true defines by $p(\boldsymbol{\mathcal{B}}|\boldsymbol{X}, \Omega_i)$, given the observations of $\boldsymbol{X}$ and Ω. The bandpass-filtered signals $\boldsymbol{\mathcal{Z}}$ are obtained by applying frequency band $\boldsymbol{\mathcal{B}}$ in raw EEG signals $\boldsymbol{X}$. By replacing the EEGs $\boldsymbol{X}$ with the bandpass-filtered signals $\boldsymbol{\mathcal{Z}}$ as follows

$$p(\boldsymbol{\mathcal{B}}|\boldsymbol{\mathcal{Z}}_i, \Omega_i) = \frac{p(\boldsymbol{\mathcal{Z}}_i, \Omega_i|\boldsymbol{\mathcal{B}})p(\boldsymbol{\mathcal{B}})}{p(\boldsymbol{\mathcal{Z}}_i, \Omega_i)}. \tag{19}$$

Moreover, particle-based approximation techniques are utilized to address the complexity problem of $p(\boldsymbol{X}, \Omega_i|\boldsymbol{\mathcal{B}})$ in Eq. (18) which is resulting in complex $p(\boldsymbol{\mathcal{B}}|\boldsymbol{X}, \Omega_i)$. The Eq. (19) is rewritten as a follow which can provide all the information regarding $\boldsymbol{\mathcal{B}}$ obtained from the bandpass-filtered signal $\boldsymbol{\mathcal{Z}}_i(k)$ and its corresponding class label Ω_i

$$p(\mathbb{B}(k)|\boldsymbol{\mathcal{Z}}_i(k), \Omega_i) = \frac{p(\boldsymbol{\mathcal{Z}}_i(k), \Omega_i|\mathbb{B}(k))p(\mathbb{B}(k))}{p(\boldsymbol{\mathcal{Z}}_i(k), \Omega_i)}. \tag{20}$$

In this method, a set of N_p particles $\{\mathbb{B}(k)\}_{k=1}^{N_p}$ generated from the prior density $p(\boldsymbol{\mathcal{B}})$ are utilized Grenander, Chow, and Keenan [2012]. $\mathbb{B}(k)$ denotes a particle representing a single filter bank. In particular, each particle contains the weight of itself ($\pi(k)$) and the characteristics of the spectral filter that it defines ($\mathbb{B}(k) = \{b_s(k), b_e(k), \pi(k)\}$).

Spatial Filtering: $\boldsymbol{W}$ Eq. (7) is found from $\boldsymbol{\mathcal{Z}}$ via a CSP algorithm Blankertz et al. [2008], then a feature vector is extracted by computing simple matrix multiplication between $\boldsymbol{\mathcal{Z}}$ and $\boldsymbol{W}$. In other words, the posterior $p(\boldsymbol{\mathcal{B}}|\boldsymbol{\mathcal{Z}}, \Omega_i)$ can be estimated from $p(\boldsymbol{F}|\boldsymbol{\mathcal{Z}}, \Omega_i)$, where $\boldsymbol{F} = log[var(\boldsymbol{W}^\dagger \boldsymbol{\mathcal{Z}})]$. A feature matrix which contains all the features for all the trials is formed as follows

$$\boldsymbol{F}(k) = \{\boldsymbol{f}_i(k)\}_{i=1}^{N_{\text{Trial}}} \in \mathbb{R}^{2m \times N_{\text{Trial}}}. \tag{21}$$

Then the posterior probability is modified as

$$\begin{aligned} p(\mathbb{B}(k)|\boldsymbol{\mathcal{Z}}_i(k), \Omega_i) &\triangleq p(\mathbb{B}(k)|\boldsymbol{F}(k), \boldsymbol{\Omega}) \\ &= \frac{p(\boldsymbol{F}(k), \boldsymbol{\Omega}|\mathbb{B})p(\mathbb{B}(k))}{p(\boldsymbol{F}(k), \boldsymbol{\Omega})}. \end{aligned} \tag{22}$$

Then, the mutual information framework is deployed regarding computing the likelihood probability

Algorithm 2 Bayesian Spatio-spectral Filter Optimization Step by Step

Input: A set of training data $\{\boldsymbol{X}, \Omega\}$

- $\{\boldsymbol{X}\}_{i=1}^{N_{\textbf{Trial}}}$: Available set of EEG epochs (trials)
- $\{\Omega\}_{i=1}^{N_{\textbf{Trial}}}$: Corresponding class labels
- N_p: Number of particles used for posterior pdf estimation.

Output: The optimized particles $\{\mathbb{B}(k)\}_{k=1}^{N_p}$ and N_p number of trained classifiers.

- Define N_p number of particles.
- Initialize the filter bank band limits with a random value.

  For the number of iterations

    For the number of particles

      Spectrally filtering the signals

      Deriving the CSP filters for each frequency band

      Spatially filtering the signals

      Extracting features

    End For

  Form the features matrix F extracting from each particle

  Computing the posterior probability and Calculate the particle weights

  Calculating the particle weights

  End For

- Train N_p number of SVM classifiers based on the features F(k) from the last iteration

$p(\boldsymbol{F}(k), \boldsymbol{\Omega})$ and measure the discriminative power of the features, where the mutual information between the features matrix and the class labels is donated by $I(.)$

$$p(\boldsymbol{F}(k), \boldsymbol{\Omega}|\mathbb{B}) = \exp\Big[I\big(\boldsymbol{F}(k); \boldsymbol{\Omega}\big)\Big], \tag{23}$$

Now, it is necessary to compute the weight of the particles, where $p(\boldsymbol{F}(k)$ denotes a feature vector set extracted from the spectrally $\mathbb{B}(k)$ and spatially $\boldsymbol{W}(k)$ filtered signals and $p(\boldsymbol{F}(k), \boldsymbol{\Omega}|\mathbb{B}_{\mathbf{k}})$

denotes the conditional observation density.

$$\pi(k) = \frac{p(\boldsymbol{F}(k), \boldsymbol{\Omega}|\mathbb{B}_{\mathbf{k}})}{\sum_{k=1}^{N_p} p(\boldsymbol{F}(k), \boldsymbol{\Omega}|\mathbb{B}(\mathbf{k}))}. \tag{24}$$

Based on Eq. (24) the set of particle weights are calculated Mohammadi and Asif [2013] and is formed as $\boldsymbol{\Pi} = \bigcup_k (\pi(k) > \tau)$, where τ is a random number between 0 and 1 and is generated at each iteration. From a Bayesian point of view, the effective prior for each iteration should be derived from the output of the previous iteration. In other words, in the tth iteration, first task is to choose a particle b_k^t with a probability π_k^{t-1} with replacement from $\mathbb{B}^{t-1}$ and operation is repeated K times resulting the new particle set $\mathbb{B}^{(t)}$. The main advantage of this particle-based approximation of the posterior density is that we can naturally obtain a data-driven filter bank that is composed of multiple particles, each of which may have a different weight and bandwidth and can possibly be overlapped. In the last step, based on the features extracted from each particle N_p number of support vector machines (SVMs) classifiers Burges [1998] are trained.

In the training phase, N_p number of optimized particles are achieved and N_p different spectral filters are applied to the trials following by applying spatial filters corresponding to each particle aiming to obtain the features of each particle. then, the output of each classifier is multiplied by the weight of the particles, and the decision rule in this method is

$$\lambda_{test} = \text{sign}\big(\sum_{k=1}^{N_p} \pi(k)\zeta(k)(\boldsymbol{X}_{test})\big). \tag{25}$$

Where, $\{\zeta(k)\}_{k=1}^{N_p}$ named the classifiers set for all the particles.

2.2.6 Double-band Bayesian Spatio-spectral Filter Optimization (B2B-SSFO)

In order to determine the optimized spectral filters to better discriminate the EEG trials, Reference Shahtalebi and Mohammadi [2018] proposed the *Bayesian Double Band Spectro-spatial Filter Optimization (B2B-SSFO)*, which aims to enhance the classification accuracy by simultaneously taking advantage of filterbank solutions and optimized Bayesian mechanisms. The B2B framework

consists of a subject-specific spectral filtering algorithm with two optimized spectral filters. The reason behind selecting two frequency bands is the fact that the motor-related features of the EEG signals are mainly stored in μ (8-13 Hz) and β (13-30 Hz) bands.

In the B2B framework, a filter bank consisting of 2 bandpass filters is randomly initialized and then the frequency limits are optimized. Similar to the discussion in Section 2.2.5, the uncertainty in the cut-off frequencies of the spectral filters are modeled with *a priori* probability denoted by $p(\boldsymbol{\mathcal{B}}^{\mathcal{D}})$ over the reference random variable $\boldsymbol{\mathcal{B}}^{\mathcal{D}}$. Unlike Suk and Lee [2013], which defines $\boldsymbol{\mathcal{B}}^{\mathcal{D}} = [b_s, b_e]$ as the cutoff frequencies of a bandpass filter, it is modified and defined as follows

$$\boldsymbol{\mathcal{B}}^{\mathcal{D}} \triangleq [b_s, b_m, b_e], \tag{26}$$

where $[b_s, b_m]$ defines the first bandpass filter, which extracts μ band contents. Band $[b_m, b_e]$ represents a second bandpass filter used to extract information from the β band. Similar to the BSSFO framework in Eq. (20) the posterior distribution denoted by $p(\boldsymbol{\mathcal{B}}^{\mathcal{D}}|\boldsymbol{X}_i, \Omega_i)$ is constructed from single-trial EEG recording $\boldsymbol{X}_i$ for $(1 \leq i \leq N_{\text{Trial}})$ and its particle-based approximation is computed as

$$p(\mathbb{B}^{\mathcal{D}}(k)|\boldsymbol{X}_i, \Omega_i) = \frac{p(\boldsymbol{X}_i, \Omega_i|\mathbb{B}^{\mathcal{D}}(k))p(\mathbb{B}^{\mathcal{D}}(k))}{p(\boldsymbol{X}_i, \Omega_i)}, \tag{27}$$

where $(\mathbb{B}^{\mathcal{D}}(k) = \{b_s(k), b_m(k), b_e(k), \pi(k)\})$, $(1 \leq k \leq N_p)$, donates particle set k and contains particles weight $\pi(k)$. It represents a single filter bank among the N_p instances used. Notation $l = 1$ represents filter $[b_s, b_m]$, while $l = 2$ represents filter $[b_m, b_e]$. The output $\boldsymbol{\mathcal{Z}}_l$ of each bandpass filter is modeled by the convolution of the input signal $\boldsymbol{X}_i$ for $(1 \leq i \leq N_{\text{Trial}})$ with system $h_l(k)$, $l \in \{1, 2\}$, and is given by

$$\boldsymbol{\mathcal{Z}}_{i,l}(k) = h_l(k) \circledast \boldsymbol{X}_i, \quad l \in \{1, 2\}. \tag{28}$$

The likelihood and evidence are, respectively, defined as $p(\boldsymbol{\mathcal{Z}}_{i,l}(k), \Omega_i|\boldsymbol{\mathcal{B}}^{\mathcal{D}}(k))$ and $p(\boldsymbol{\mathcal{Z}}_{i,l}(k), \Omega_i)$ by replacing the raw EEG signal $\boldsymbol{X}_i$ with its bandpass-filtered version $\boldsymbol{\mathcal{Z}}_{i,l}(k)$ as

$$p(\mathbb{B}^{\mathcal{D}}(k)|\boldsymbol{\mathcal{Z}}_{i,l}(k), \Omega_i) = \frac{p(\boldsymbol{\mathcal{Z}}_{i,l}(k), \Omega_i|\mathbb{B}^{\mathcal{D}}(k))p(\mathbb{B}^{\mathcal{D}}(k))}{p(\boldsymbol{\mathcal{Z}}_{i,l}(k), \Omega_i)}. \tag{29}$$

The spectral filtering step is followed by computing CSP features of each iteration for each frequency band (l) in each particle, similar to Eq. (7), i.e.,

$$\boldsymbol{\mathcal{W}}_{i,l}(k) = \big[\boldsymbol{W}_l(k)\big]^T \boldsymbol{Z}_{i,l}(k). \tag{30}$$

Based on Eq. (30), the corresponding features for the trials in each frequency band are extracted as

$$\boldsymbol{f}_{i,l}(k) = \log\left(\frac{\text{var}\big(\boldsymbol{\mathcal{W}}_{i,l,p}(k)\big)}{\sum\limits_{i}\text{var}\big(\boldsymbol{\mathcal{W}}_{i,l,p}(k)\big)}\right). \tag{31}$$

Next, the features of the two parallel pipelines are merged into one single feature vector as

$$\boldsymbol{f}_i(k) = \Big[\boldsymbol{f}_{i,l}(k)|_{l=1}, \boldsymbol{f}_{i,l}(k)|_{l=2}\Big], \tag{32}$$

The rest of the procedure is as the same as Section 2.2.5. This completes my overview of all the necessary concepts of understanding BCIs and different methods in feature extraction step before I move on to describing my contributions. Next two chapters describe the frameworks I proposed and investigated during my research work.

Chapter 3

Experimental EEG-based Benchmark Data Collection

The objective of this chapter is to focus on the element of experimental design and construct a benchmark dataset to further advance EEG signal processing techniques. In other words, to further examine the effects of deploying optimized subject-specific spectra-spatial filters, it is vital to examine and investigate different aspects of data collection and in particular, effects of the stimuli provided to a subject to trigger actual motion or MI tasks. Data collection is performed through a portable and wireless biosignal acquisition system, g.Nautilus from g.tech Medical Engineering g.Tech: g.Nautilus [2018]. The EEG signals are recorded from 10 healthy volunteer subjects performing *Actual Movement (AM)* and *Motor Imagery (MI)* tasks. The subjects are asked to perform either left-hand movement, referred to as Class $C1$, or right-hand movement, referred to as Class $C2$, depending on a visually presented cue. We performed several experiments on the same subjects with four different protocols of stimulations with visual and voice stimulus, which will be described later in this chapter. Moreover, for validating the collected data, the CSP method described in Chapter 2 is utilized and results are presented in this chapter.

The rest of the chapter is organized as follows: In Section 3.1, the acquisition equipment utilized for data collection is introduced together with the accompanying hardware and software components. Section 3.2 focuses on description of experimental design paradigm, where different design

Figure 3.1: g.Nautilus Research Headset.

scenarios for recording signal. Finally, Section 3.3 presents the entire process to interpret and evaluate the collected data.

3.1 EEG Signal Recording

In this section, the acquisition equipment utilized for constructing the benchmark EEG dataset is introduced together with the required signal recording components. Different aspects of the utilized EEG is outlined below:

- ***g.Nautilus Research EEG system***: The acquisition equipment used for data collection, which is a wireless biosignal acquisition system g.Tech: g.Nautilus [2018], is designed by g.Nautilus and can collect EEG data with 24 Bit resolution and a sampling rate of 250 Hz or 500 Hz. The g.Nautilus headset is a wireless biopotential amplifier with prefixed electrode strands and a cap with 32 active bipolar electrodes, as shown in Fig.3.1. Each electrode side is labeled with

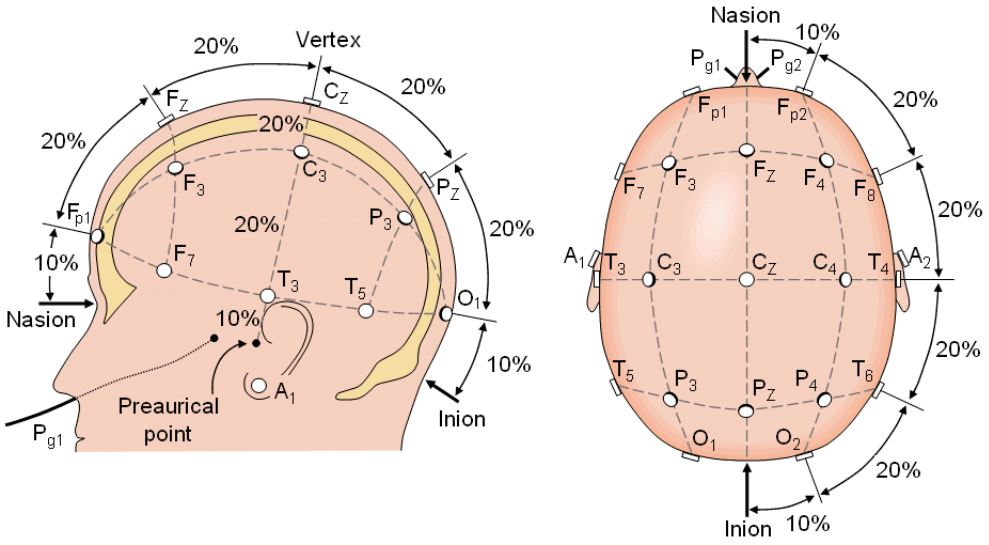

Figure 3.2: An example of 10-20 setting of EEG electrodes placement. electroencephalography book [2018] .

a number and a letter. Letters are assigned to different brain area, e.g., Letter "F" refers to the Frontal lobe, while Letter "T" refers to the Temporal lobe. Numbers stand for either right or left side of the brain, i.e., even numbers denote the right side of the head and odd numbers the left side of the head.

- ***Electrode Positioning***: The electrodes are uniformly distributed according to the special positions, which are identified using the international 10-20 system (Fig. 3.2). The basis of this system is the distance in percentage scale between Nasion-Inion and fixed points marked as the Frontal pole (Fp), Central (C), Parietal (p), Occipital (O), and Temporal(T). Subscript "Z" is used to refer to the mid-line electrodes symbolized for Zero.

- ***Electrode Gel***: For reducing the resistance between each electrode and the scalp to less than 5 kOhm, which is the optimal value for EEG recording, the electrodes are filled with conductive gel. Especially, using gel maximizes skin contact and allowing low-resistance recordings through the skin.

3.1.1 Software Required to record the EEG signal

The following two software packages are required in the data collection setup:

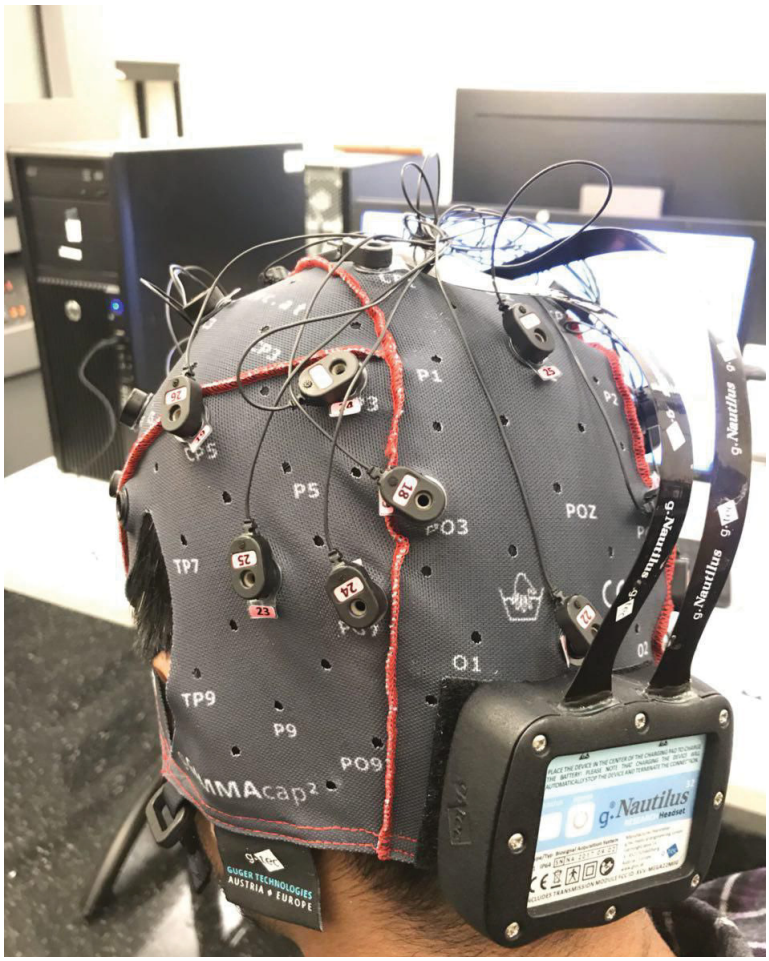

Figure 3.3: g.Nautilus Research Headset, a wireless biosignal acquisition system.

- *The Network Enabled Easy Data Access (g.NEED access)*, which is a server service that comes with the application for displaying the raw EEG data. Besides, it can be used for basic testing, such as channel impedance measurement. Basically, this test easily indicates the impedance of connected electrodes.

- *g.NEEDaccess MATLAB API*, which is a MATLAB script that requires the installation of g.NEEDaccess and allows us to set up the configuration and collect EEG signal.

3.1.2 Setting Up the EEG Headset

First of all, the EEG cap (Fig. 3.3) is placed on the head of a subject aligning the electrodes according to the international 10-20 electrode system. Fig 3.2 is an example of a 10-20 EEG electrode

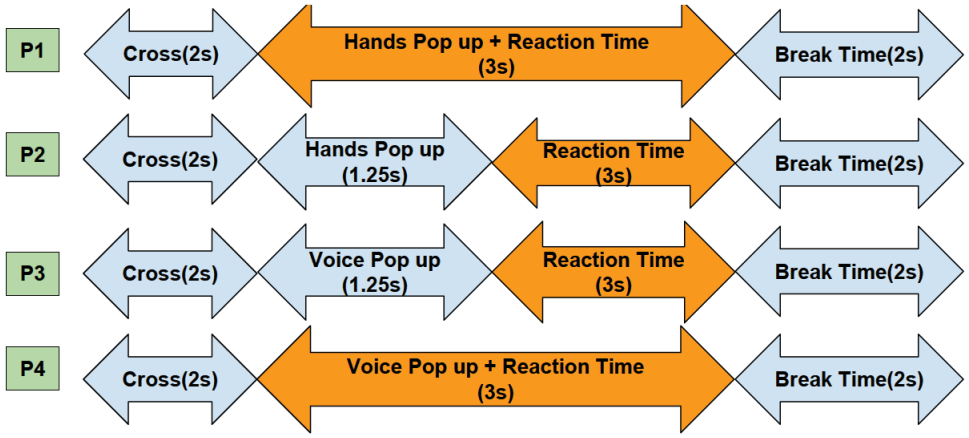

Figure 3.4: Timing schemes for different protocols P1, P2, P3, P4 used to collect EEG signals: P1 and P2 are stimulated visually by presenting information on a screen, and; P3 and P4 are based on voice stimulation.

placement system. The conductive gel is used to fill into the electrodes completely. After connecting base station, using the cable to a USB port on the PC, the device is switched on by pressing the power button. After a short time, the status-LED of the headset blink slowly, and the headset connects to the base station. Moreover, after setting up the headset, g.NEEDaccess is used to check the impedance of each electrode. Larger numbers mean higher resistance to current flow, therefore, the conductive gel would be used again for those channels with high impedance to bring it down below 15kOhm.

3.2 Experimental Paradigm and EEG Recording

Different protocols, as shown in Fig. 3.4, are conducted where the subjects are asked to perform repetitively either left/Right hand movement or imagine about doing these movements. As shown in Fig. 3.5, for each protocol, i.e., P1, P2, P3, and P4, a recorded trial starts with the presentation of a fixation cross in the center of screens, which followed by four different paradigms and randomized breaks between consecutive trials (2 or 3 seconds). In three experimental runs of 100 trials, each subject was asked to perform actual movements of the left or right hand or imagine doing these movements. The total duration of the reaction time for each trial is 3 seconds. The stimuli displayed on the screen corresponds to one of the following four different protocols:

- ***Protocol P1***: Starts with a 2-second initialization with a cross shown on the screen, followed

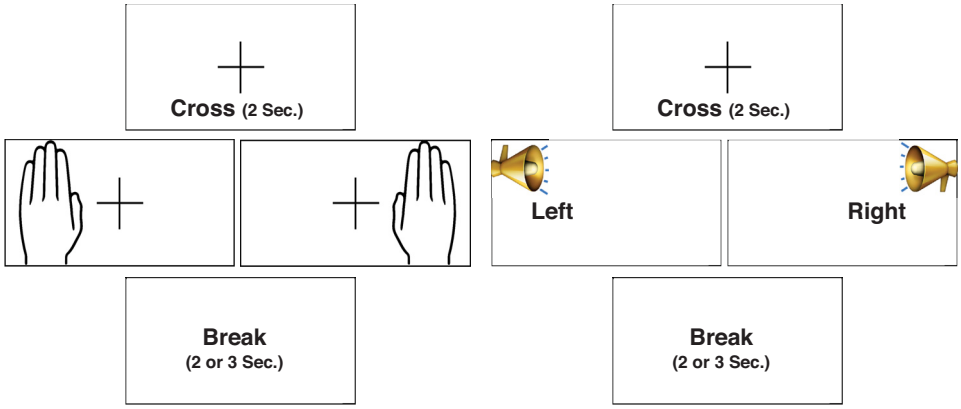

Figure 3.5: Experimental Procedure Starting with the fixation cross in the screen, followed by either popping up hands or speech and randomized breaks between consecutive trials.

by a left or right hand stimulus maintained for 3 seconds while the brain activity is being simultaneously monitored. There is a 2-second break before the introduction of the next trial.

- ***Protocol P2***: Similar to Protocol P1, except that the visual stimulus is popped up for 1.25 seconds. After which, the brain activity is monitored for 3 seconds with the subject performing a given task.

- ***Protocol P3***: Unlike Protocols P1 and P2, subjects keep their eyes closed in the trial. The stimulus is provided orally. The 3-second EEG recording starts after 1.25 seconds of the voice stimulation.

- ***Protocol P4***: Similar to Protocol P3, except that the EEG is recorded simultaneously while the voice stimulus is maintained.

3.2.1 Ethical Issues

Experimentation on animal and human subjects has a long history in the scientific and medical literature. Research risks are always there and unavoidable. Therefore, the use of human subjects in experimentation, especially in health care generates ethical, legal, political, and humanistic concerns. The data was collected with the policy certification of Ethical acceptability for research involving human subjects and approved by Concordia University with the certification

number 30007997. The procedures used in this protocol are all well established and known to be safe.

3.2.2 Data Recording

Dataset has been acquired from 10 healthy volunteer subjects during eight measurement sessions. All individuals were sitting in a standard chair, relaxed and watching the screen placed in the way of their eyes. Each subject performed the tasks associated with each protocol twice, one performing an actual movement (AM) and one performing motor imagery (MI). Depending on the presented cue, the subjects are asked to perform repetitively either a left-hand movement, referred to as Class $C1$, or a right-hand movement, referred to as Class $C2$ during different designed protocols. The brain activities were recorded by 32 bipolar channels with a sampling frequency rate of 500 Hz.

As stated previously, for each subject, eight data collection sessions are conducted, and each session consists of three experimental runs of 100 trials. Each trial is started with the presentation of a fixation cross in the centre of the screen, followed by four different paradigms and randomize breaks between consecutive trials (ranging between 2 to 3 seconds). The total duration of reaction time for each protocol is 3 seconds.

To ensure a synchronized and unified measurement of signals in practical experiments, a synchronizing device is used. This device enables the user to have real-time access to raw signals and the option for centralized collection, if required, when more than two devices are involved in data collection experiments. Synchronizing device is a Light Dependent Resistor (LDR) in a voltage divider circuit, which increases resistance in the dark and decreases under the light. To implement the circuit, a 4.5-volt battery pack and a 10kOhm resistor were connected in series with the photocell. Finally, the output was taken from the 10kOhm resistor and connected to the digital input pins on the g.Nautilus Research base station to convert the high and low analog voltage signal to 1 second and 0 second digital signals acquired by the base station along with the EEG signals from the subject.

In summary, the brain activities were recorded in trials with a total duration of 9.25 seconds, including 1 sec preparation where a white cross is displayed in the middle of the screen; 1.25 sec for presentation of the stimulus; 4 sec where an empty screen is displayed for the subject to imagine

or perform the movement, and; finally, 2 or 3 sec break representing black screen to avoid the brain adaptation problem. EEG data is recorded in 3 sections, each including 100 trials and the 4-minutes long break between them.

3.2.3 Data File Description

Each participant appeared in 8 data collection sessions, within each, different protocols were instructed to the subject. The collected data was gathered from 10 healthy subjects performing AM and MI tasks. The collected data, hereafter referred to as "dataset", includes two .MAT files, i.e., "Class-Labels-Dataset", which is a 2-dimensional matrix ($1\times$ number of trials), and; "Raw-Signal-Dataset", which is a 3-dimensional matrix. Within the Raw-Signal-Dataset, the first dimension refers to the number of the electrodes (32 bipolar electrodes); The second dimension shows the duration of the participant's reaction (3 seconds or 1500 Time Samples, and; The third dimension refers to the number of trial (i.e., 300 trials).

3.3 Requirement and Evaluation and Results

First and foremost, the performance of a BCIs system is evaluated via its classification accuracy. In terms of evaluating the collected datasets, the outcome of the classifier is then compared with the ground truth provided in the labels of the dataset.

Collected raw EEG signals from the surface of the scalp are not necessarily an accurate representation; they tend to contain multiple signals from different sources such as, artifacts, e.g., blinking or muscle movement, noise, and brain signals which can obscure weaker EEG signals. *Preprocessing step* can be considered as the first vital block before the processing methods, aid in improving the performance of the system by separating the noise from the actual signals. In other words, Preprocessing techniques help to intensify the signals and upgrade signal to noise ratio (SNR) Yang, He, Wang, Song, and Zhang [2014]. The general step is applying a notch filter centred at 60Hz to remove the power line interference. Chebyshev-II filter of order 10, which is a Bandpass filter, is applied to get the desired band of frequency contents within 0.5 to 100 Hz Ramoser, Muller-Gerking, and Pfurtscheller [2000]. Finally, the signals are filtered using the Chebyshev-II filter of order 20

within 4 to 40 Hz.

Regarding the pre-processing step, the following two approaches are initially studied: (i) The first approach was to use all the recorded signal at once, apply bandpass filtering, and subsequently divide the filtered data into epochs, and; (ii) The second approach was to first identify epochs and then apply bandpass filtering on the extracted ones. Both approaches were tested and compared for accuracy. Based on the evaluation results, the first approach was identified as the better solution and considered in our work.

Feature extraction step is done after the preprocessing step aiming to extract the relevant information corresponding to different mental activities. Common Spatial Patterns (CSP) Townsend, Graimann, and Pfurtscheller [2006] is employed as a feature extraction procedure to derive spatial ERD/ERS features Müller-Gerking, Pfurtscheller, and Flyvbjerg [1999]. Feature extraction combining with selection of a set of *Optimal Hyper-parameters* can significantly improve the model's performance. The hyper-parameters to be optimized are as follows: (i) The length of the trimming window used to select the most informative section of the recorded signal within each 3 seconds epoch; (ii) Number of spatial filters utilized within the feature extraction step, and; (iii) The choice of the classifier. These three items are tuned via cross-validation method Kohavi et al. [1995] as described below:

- ***Trimming:*** The proposed trimming framework modifies time intervals for each epoch, which corresponds to adaptively seeking for the informative section of the signal. The start time $\boldsymbol{t}_s$ is fixed to zero, and the goal is to find a subject-specific end-time resulting in selection of the optimal information-bearing section of the signal.

- ***Selection of Eigenvectors:*** As a mathematical algorithm, the CSP method computes spatial filters that aim at achieving optimal discrimination by decomposing the eigenvalues and eigenvectors of the spatial covariance matrix for maximum variance difference between the two classification classes. The goal here is to find a subject-specific number of eigenvectors to be selected from the decomposed spatial covariance matrix

- ***Selection of the Classifier:*** Three classifiers are deployed and compared for accuracy, i.e., Linear Discriminant Analysis (LDA) Scherer et al. [2004], Quadratic Discriminant Analysis

(QDA), and linear kernel of Support Vector Machines (SVM) Khasnobish, Bhattacharyya, Konar, Tibarewala, and Nagar [2011].

For finding the best combination of hyper-parameters, the dataset is randomly shuffled ten times to avoid bias in the split datasets before training the models. The shuffled dataset is then divided into a *Train-Set*, and a *Test-Set*. We considered 70% of the dataset as the train-set, and the rest is set aside for testing. Following the cross-validation approach Wong [2015], the Train-Set is divided into $k = 10$ roughly equal sections, each unique group k considered as a holdout or validation-set and the remaining groups as the train-set. The model is fitted on the train-set and evaluated on the validation-set. The goal is to provide an opportunity for each sample to be used in the hold-out-set once and then be used to train the model $k - 1$ times. Trimming loop was applied to all the training trials, where we assumed that the minimum time interval required to respond to the stimulus and performing the requested task is 1 second within each preprocessed epoch of 3 seconds. To find the best number of feature vectors $\{\varrho_1 \geq \varrho_2 \geq \cdots \varrho_n\}$, eigenvalues $\{\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n\}$, obtained from decomposing the spatial covariance matrix, are stored in a descending order. A grid search is performed to achieve maximum performance by picking up an equal number of the eigenvectors (2, 4, or 6 eigenvectors) to build the whitening matrix. Based on the defined task for classification, the classifier that provides maximum accuracy on the validation set is selected. The results of validating datasets belonging to different subjects are shown Tables 3.1-3.4.

Table 3.1: *Protocol P1* in *Actual Movement (AM))* and *(Motor Imagery (MI)*: Average performance of the cross-validation method using conventional CSP method based on different tuning parameters for *10 Subjects*.

Protocol 1 AM	**0-150**	**0-300**	**0-450**	**0-600**	**0-750**	**0-900**	**0-1050**	**0-1200**	**0-1350**	**0-1500**
LDA - 2Eig	71.42	67.70	64.85	62.67	62.20	61.98	62.56	61.42	61.29	61.65
LDA - 4Eig	74.71	70.71	67.89	66.82	67.06	67.12	67.43	68.24	68.59	68.32
LDA - 6Eig	75.29	71.74	68.85	68.63	69.19	70.01	70.18	70.95	69.41	69.44
QDA - 2Eig	69.99	66.46	63.69	62.31	60.69	60.58	61.05	60.05	59.38	60.06
QDA - 4Eig	74.15	69.35	66.26	65.51	65.09	65.53	65.57	66.71	66.80	67.08
QDA - 6Eig	74.30	69.95	67.67	67.37	66.57	67.81	68.31	69.28	69.39	69.49
SVM - 2Eig	70.95	67.27	64.63	62.59	62.06	61.86	61.66	61.93	61.11	61.81
SVM - 4Eig	74.59	70.56	67.93	66.80	67.13	67.52	67.73	68.52	68.82	69.06
SVM - 6Eig	75.05	71.78	69.20	68.59	68.74	69.74	70.57	71.42	**71.61**	69.50

Protocol 1 MI	**0-150**	**0-300**	**0-450**	**0-600**	**0-750**	**0-900**	**0-1050**	**0-1200**	**0-1350**	**0-1500**
LDA - 2Eig	66.15	60.77	59.58	59.08	59.27	58.47	57.76	57.01	55.95	54.94
LDA - 4Eig	69.30	65.34	63.88	63.72	63.41	61.41	61.18	61.23	60.48	59.30
LDA - 6Eig	69.34	66.03	64.97	64.67	64.91	61.53	61.49	61.46	61.48	59.40
QDA - 2Eig	64.81	59.00	58.38	58.43	57.77	57.49	56.33	55.70	54.87	54.16
QDA - 4Eig	68.24	63.74	62.80	61.70	61.63	60.83	59.28	59.29	58.64	57.32
QDA - 6Eig	68.56	64.39	63.11	62.42	62.10	61.47	60.37	60.36	59.92	59.05
SVM - 2Eig	65.67	60.63	59.30	59.01	59.02	57.80	57.44	56.45	55.38	54.59
SVM - 4Eig	69.12	65.37	64.15	63.63	63.83	62.34	61.17	60.99	60.15	59.24
SVM - 6Eig	69.37	65.88	64.73	64.76	64.80	61.57	**61.55**	61.36	61.33	59.54

Table 3.2: *Protocol P2* in *Actual Movement (AM))* and *(Motor Imagery (MI)*: Average performance of the cross-validation method using conventional CSP method based on different tuning parameters for *10 Subjects*.

Protocol 2 AM	**0-150**	**0-300**	**0-450**	**0-600**	**0-750**	**0-900**	**0-1050**	**0-1200**	**0-1350**	**0-1500**
LDA - 2Eig	57.55	55.86	56.36	57.66	59.41	59.86	59.86	60.77	60.84	60.90
LDA - 4Eig	57.91	57.13	58.02	59.26	61.09	61.69	62.41	62.91	63.91	64.11
LDA - 6Eig	58.54	58.23	59.65	61.50	63.75	64.28	65.69	66.84	68.22	65.30
QDA - 2Eig	55.82	54.45	54.02	55.82	56.42	57.05	56.99	57.41	58.35	57.60
QDA - 4Eig	57.21	56.03	55.66	57.23	58.53	59.95	59.79	60.48	61.33	61.23
QDA - 6Eig	57.54	56.97	58.02	59.76	60.96	62.09	63.14	64.33	64.70	65.28
SVM - 2Eig	56.58	55.56	55.17	56.99	58.91	59.11	59.00	60.16	60.52	60.42
SVM - 4Eig	58.91	57.93	58.47	59.71	62.11	62.78	62.91	64.17	64.91	65.02
SVM - 6Eig	58.35	58.33	59.38	61.12	63.44	64.16	64.91	66.35	**69.29**	65.45

Protocol 2 MI	**0-150**	**0-300**	**0-450**	**0-600**	**0-750**	**0-900**	**0-1050**	**0-1200**	**0-1350**	**0-1500**
LDA - 2Eig	53.68	55.37	56.33	56.48	58.66	58.54	59.06	58.76	58.52	57.67
LDA - 4Eig	54.27	56.18	58.01	58.91	61.99	61.77	62.71	63.10	62.87	60.88
LDA - 6Eig	54.64	56.58	58.92	60.30	62.55	63.41	63.98	64.23	64.17	61.48
QDA - 2Eig	52.57	54.81	55.72	56.09	56.77	57.31	58.21	57.89	58.31	58.12
QDA - 4Eig	53.88	54.65	56.28	57.31	59.89	60.51	61.43	61.36	61.48	60.00
QDA - 6Eig	54.31	55.39	57.09	58.36	60.72	61.45	61.99	62.36	62.74	61.07
SVM - 2Eig	53.66	54.94	56.19	56.70	57.82	58.92	58.72	58.33	58.24	57.76
SVM - 4Eig	54.53	55.88	58.19	59.25	62.05	62.04	62.64	63.22	63.18	61.06
SVM - 6Eig	54.48	56.58	58.68	59.96	62.92	63.35	63.99	**64.35**	64.20	61.52

Table 3.3: *Protocol P3* in *Actual Movement (AM))* and *(Motor Imagery (MI)*: Average performance of the cross-validation method using conventional CSP method based on different tuning parameters for *10 Subjects*.

Protocol 3 AM	**0-150**	**0-300**	**0-450**	**0-600**	**0-750**	**0-900**	**0-1050**	**0-1200**	**0-1350**	**0-1500**
LDA - 2Eig	57.85	62.29	64.65	67.89	68.04	69.18	68.76	68.35	69.29	70.72
LDA - 4Eig	60.20	63.93	67.14	70.55	70.90	71.97	72.82	72.39	73.46	71.26
LDA - 6Eig	60.70	64.88	67.53	70.94	71.31	72.51	73.72	74.25	75.25	71.62
QDA - 2Eig	58.05	61.72	64.36	67.95	67.58	69.27	68.72	66.99	68.97	70.05
QDA - 4Eig	60.42	63.45	66.74	69.51	69.94	70.93	71.60	70.91	71.85	71.93
QDA - 6Eig	60.14	63.97	66.87	69.58	70.03	71.38	72.30	72.77	73.45	72.11
SVM - 2Eig	57.78	61.69	64.38	67.89	68.38	69.55	69.15	68.71	69.76	71.00
SVM - 4Eig	59.81	64.28	67.46	70.37	71.03	72.16	72.83	72.40	73.77	71.60
SVM - 6Eig	60.84	64.77	67.43	70.82	71.18	72.47	73.56	74.38	**75.57**	72.69

Protocol 3 MI	**0-150**	**0-300**	**0-450**	**0-600**	**0-750**	**0-900**	**0-1050**	**0-1200**	**0-1350**	**0-1500**
LDA - 2Eig	51.84	51.34	52.16	52.30	52.46	51.51	51.63	52.06	52.35	52.50
LDA - 4Eig	51.75	52.02	52.71	53.72	54.72	53.75	54.58	54.90	55.34	55.46
LDA - 6Eig	53.11	52.82	53.74	54.38	54.77	54.07	54.96	55.71	56.51	55.90
QDA - 2Eig	50.87	50.93	51.38	51.28	51.37	50.87	51.00	51.34	51.85	51.39
QDA - 4Eig	51.45	51.34	52.14	52.20	53.31	52.47	53.32	53.46	54.37	54.28
QDA - 6Eig	52.00	51.94	52.78	53.02	53.92	53.21	53.64	54.71	56.08	54.92
SVM - 2Eig	51.82	51.18	52.20	51.99	52.09	51.20	51.35	51.79	52.14	52.50
SVM - 4Eig	51.55	51.77	52.89	53.63	54.17	53.83	54.02	54.39	56.26	55.25
SVM - 6Eig	52.89	52.57	53.90	54.70	55.03	54.50	55.01	55.42	**57.67**	56.28

Table 3.4: *Protocol P4* in *Actual Movement (AM))* and *(Motor Imagery (MI)*: Average performance of the cross-validation method using conventional CSP method based on different tuning parameters for *10 Subjects*.

Protocol 4 AM	**0-150**	**0-300**	**0-450**	**0-600**	**0-750**	**0-900**	**0-1050**	**0-1200**	**0-1350**	**0-1500**
LDA - 2Eig	53.27	50.99	50.29	50.05	50.79	51.97	54.22	54.79	55.53	56.27
LDA - 4Eig	54.00	51.40	51.72	53.42	55.88	53.36	52.15	53.86	55.13	56.18
LDA - 6Eig	53.57	51.95	52.46	55.23	54.13	51.39	56.05	55.65	58.62	59.90
QDA - 2Eig	53.56	50.91	50.15	50.29	51.00	51.46	53.11	53.31	54.47	55.33
QDA - 4Eig	53.70	51.42	51.98	52.98	54.93	52.42	50.08	50.31	52.78	54.41
QDA - 6Eig	53.73	51.87	52.20	56.47	54.71	52.09	52.93	52.89	56.02	58.67
SVM - 2Eig	53.48	50.63	50.12	49.81	50.48	52.02	53.95	53.93	56.07	55.86
SVM - 4Eig	53.85	51.83	51.75	53.46	53.29	51.85	55.96	56.29	59.78	56.42
SVM - 6Eig	53.86	52.16	52.59	55.61	52.18	58.57	**62.53**	62.15	61.40	61.52

Protocol 4 MI	**0-150**	**0-300**	**0-450**	**0-600**	**0-750**	**0-900**	**0-1050**	**0-1200**	**0-1350**	**0-1500**
LDA - 2Eig	48.79	48.33	48.62	49.19	50.20	50.66	50.75	51.15	51.49	51.11
LDA - 4Eig	50.25	49.77	50.00	51.88	53.63	53.10	53.98	55.13	54.02	52.80
LDA - 6Eig	51.10	50.14	50.12	52.43	53.69	53.94	53.12	54.24	54.24	53.20
QDA - 2Eig	49.59	49.17	49.45	50.08	50.18	50.38	50.46	50.38	50.61	50.41
QDA - 4Eig	49.94	49.85	50.06	51.08	51.67	51.36	51.79	52.44	52.75	52.76
QDA - 6Eig	50.61	50.17	50.42	51.99	52.27	52.13	53.19	53.80	53.94	52.37
SVM - 2Eig	48.51	48.56	48.25	49.41	49.86	50.06	50.11	50.58	50.81	51.03
SVM - 4Eig	49.88	49.32	49.82	51.93	53.32	53.31	52.12	53.70	54.03	52.90
SVM - 6Eig	50.90	50.22	49.64	52.82	53.74	53.68	53.98	54.33	**54.37**	53.40

Chapter 4

Regularized, Tuned, and Subject-Specific EEG-based Feature Extraction Frameworks

Despite recent advances in signal processing, machine learning, and computational technologies, our brain remains unparalleled in its information processing abilities to analyze and fuse different multi-modal, streaming signals in an adaptive and real-time fashion. This chapter seeks to address the issues related to the processing stage of motor imagery (MI) based BCI systems. In particular and in Section 4.2, the design process is introduced consisting of initiation of the algorithm parameters using cross validation pursuing the purpose of classification accuracy improvement. In Section 4.3, the adaptive epoch trimming mechanism is introduced, which seeks the best start and stop times within each epoch, and readjusts the epochs by trimming the segments before the former and after the latter. In Section 4.4, a subject-specific spatio-spectral filtering framework, referred to as the Regularized double-Band Bayesian (R-B2B) framework, is proposed that better matches the MI classification problem and provides more accurate results in comparison to its counterparts.

4.1 Introduction

A unique characteristic of the human brain is its plasticity property, i.e., the ability of neurons to modify their behavior (structure and functionality) in response to environmental diversity. This exclusive property has recently been utilized by researchers Gill et al. [2018], and Angeli et al. [2018] to recover the ability to move legs in patients paralyzed with spinal cord injuries after training them with an exoskeleton linked to their brain. The plasticity property of brain has also motivated the design of brain-computer interfaces (BCI) J. Wolpaw and Wolpaw [2012] to develop an alternative form of communication between human brain signals and the external interfacing world.

The brain wave acquisition systems used to capture neural activities of the brain for designing a BCI system are classified in the following two categories: (i) Invasive modalities, where signals are collected from electrodes implanted surgically inside patients' brains, and; (ii) Noninvasive modalities Zander and Kothe [2011] where neural activities are recorded externally from electrodes placed on patients' scalps. Electroencephalography (EEG) belonging to the latter category is a method of choice and the preferred modality for collecting brain activities due to, among other features, its non-intrusive nature, high temporal resolution, design flexibility, possibility of replacing electrodes, and lower cost as compared to other neuroimaging techniques Nicolas-Alonso and Gomez-Gil [2012], Xie et al. [2017].

Regarding the weak potentials and interference from different physiological activities, extracting the exact electrical response of the brain to the related task is challenging. As we mentioned in Chapter 2, the common spatial pattern (CSP)) Ramoser et al. [2000] is one of the most popular and effective techniques used for classification and feature extraction. The CSP applies a linear transformation to maximize the variance in one class, while variance in the other class is minimized Xie et al. [2017]. The CSP approaches continue to suffer from sensitivity to noise and adverse over-fitting effect, which has motivated the extensive research work to improve the performance of EEG-based BCI systems. In what follows, we focus on two solutions to address this issue.

4.2 Improving Accuracy of EEG-based BCIs via Tuning Mechanism

Machine learning methods aim to develop a technique that captures some element of interest from a given dataset Claesen and De Moor [2015]. Choosing a set of optimal hyperparameters for a learning algorithm can significantly affect the resulting model's performance. Cross-validation method Kohavi et al. [1995] is a re-sampling procedure used to evaluate machine learning models on a limited dataset. This statistical method is primarily used to compare and estimate the skill of model on unseen datasets. Based on the CSP feature extraction method that was introduced in Chapter 2, we extract optimal time interval (Trimming) for each epoch, the reasonable number of eigenvectors from the decomposed spatial matrix, and the best classifiers for the classification part.

4.2.1 Tuning Algorithm Parameters

The trimming design process consists of initiation of the algorithm parameters, regardless of the technique used for extracting features. The parameters, which are to be selected, are as follows:

- ***Trimming:*** The proposed trimming framework modifies time intervals for each epoch, which means adaptively seeking for the informative part of the signals. The start time $\boldsymbol{t}_s$ is fixed to zero, and the goal is obtaining the optimal time interval for each epoch.

- ***Eigenvector Selection:*** As we mentioned before, the CSP, as a mathematical algorithm, computes spatial filters that aim at achieving optimal discrimination by decomposing eigenvalues and eigenvectors of spatial covariance matrix Shahtalebi and Mohammadi [2017a]. Optimizing the number of selected eigenvectors from the decomposed spatial covariance matrix seeking to have reliable performance is the second goal of the trimming algorithm.

- ***Classifier Selection:*** Classification is a supervised learning approach that maps every possible input available in the calibration dataset to a finite set of decisions. To classify the extracted feature vectors and to define boundaries between different targets, the best classifier that achieves the maximum accuracy is selected from the following three options: Linear Discriminant Analysis (LDA) Scherer et al. [2004], Quadratic Discriminant Analysis (QDA), and Support Vector Machines (SVM) Khasnobish et al. [2011] Shahtalebi and Mohammadi

[2017b] with linear kernels.

These parameters are tuned via employing the cross-validation method. Intuitively speaking, this approach was adopted to get the optimal combination of parameters and to improve the performance of a BCI system. To evaluate the proposed framework, the algorithm applied to the collected datasets described in Chapter 3.

For implementing the trimming algorithm, the DataSet is shuffled ten times randomly to avoid any element of bias in the datasets before training the model. This step is followed by splitting the shuffled DataSet into the *Train-Set* and the *Test-Set*. We considered 70% of DataSet as the Training set, and the rest was put aside for testing. Based on the cross-validation approach Wong [2015], the Train-Set is divided into $k = 10$ roughly equal parts. Each of the $\boldsymbol{k}$ groups is considered as a hold-out or validation-set and the remaining groups are used as the Train-Set. Model is fitted on the Train-Set and evaluated on the validation-Set. The goal is given the opportunity to each sample to be used in the hold-out set one time and used to train the model *k-1* times. Trimming loop was applied to all the training trials; we assumed that the minimum time interval required to respond to a given stimuli and perform the task is 1 second or 500-time samples of the 1500 pre-processed epoch data with 3seconds length.

In order to find the optimal number of feature vectors $\{\varrho_1 \geq \varrho_2 \geq \cdots \varrho_n\}$, after decomposing spatial covariance matrix, eigenvalues $\{\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n\}$ are stored in a descending order. A grid search is performed to have the maximum performance by picking up an equal number of eigenvectors (2, 4, or 6 eigenvectors) to build the whitening matrix. Finally, based on the stated definition for the classification task, by taking to account the maximum accuracy on the validation set, the best classifier is selected. Finally, DataSet is shuffled randomly, split up to TrainSet and TestSet (TrainSet:70%, TestSet: 30%), all the predicted parameters fitted on the Train-Set and evaluated on the Test-Set to enhance the model's performance. The Grid-search step is elaborated in Algorithm 3.

Algorithm 3 K-FOLD CROSS-VALIDATION METHOD

DataSet: Matrix X_i, size:$32 \times 1500 \times 300$,
split into **TrainSet**: $32 \times 1500 \times 210$, **TestSet**: $32 \times 1500 \times 90$

K= 10 fold, Train_trial: 70%, and Test_trial: 30%
Input:TrainSet (Size: $32 \times 1500 \times 210$)

For Classifier_Type = LDA, QDA, SVM

For Eigen_Vectors = 2, 4, or 6

For Trimming_Index = 1:Timesample/Step

For RandomTest = 1:Num_of_Rand

- The CSP is applied;
- The accuracies of classification are evaluated.

End

- Optimal Parameters are capture to get the maximum model's performance.

4.2.2 Stimulation and Results for Cross-Validation Algorithm

In order to evaluate the proposed framework, conventional CSP algorithm is used with the best configuration obtained from the parameter tuning algorithm. Different protocols explained in Chapter 3 are used to compare the effect of visual or voice stimuli on the overall performance. Results are provided in Table 4.1 and Fig. 4.1. The average performance comparison is provided here based on the conventional CSP with tuned parameters, i.e., SVM classification, and number of eigenvectors set to 6 feature vectors for Protocols P1 to P4.

4.3 Improving the Performance of MI EEG-based BCIs via an Adaptive Epoch Trimming Mechanism

As stated previously, one of the most popular and commonly used techniques to satisfy the requirement for an effective and efficient BCI is Motor Imagery, which is defined as merely imagination of a limb movement, with no actual movement or peripheral (muscle) activation. However, as much as this field outlines a promising framework, the researchers dealing with MI commonly face two types of challenges. The first challenge is to deal with *different comprehension of subjects from "imagination of the movement"*. Some subjects imagine repeating the movement during each epoch, while some others might execute the mental imagination of the activity only once, and not necessarily consistently. In order to tackle this obstacle of various reactions, implementation of

Table 4.1: *Protocols P1 to P4* for *Actual Movement (AM)* and *Motor Imagery (MI)*: Average classification performance for 10 subjects with tuned/fixed TimeSample, classification (SVM), and number of eigenvectors (6).

Avg. Accuracy of 10 Subjects	Tasks	CSP 1500 TimeSample	CSP Tuned
Protocol 1	AM	69%	71%
Protocol 2	AM	65%	69%
Protocol 3	AM	73%	75%
Protocol 4	AM	61%	63%
Protocol 1	MI	59%	61%
Protocol 2	MI	61%	64%
Protocol 3	MI	56%	58%
Protocol 4	MI	53%	54%

methods in which the classifier is trained subject by subject has been adopted by the researchers of the field. These approaches are adaptive to the nature of the datasets collected, for instance, should the subject react to the stimulus right after he/she sees it, the classifier of the dataset collected from this subject is trained to read the epochs' information right after the marker. The second challenge is the fact that through a cognitive process, the brain of the subject learns to decrease the motor concentration while doing the same task. Hence, the amplitude of the signals within each epoch tends to descend over time. In order to tackle the first challenge, one approach is implementation of subject-specific methods, i.e., an adaptive classifier is trained subject by subject. In this case, the processing framework is adaptive to the nature of the datasets collected and is robust by design. For instance, should the subject react to the stimulus right after he/she sees it, the classifier of the dataset collected from this subject is trained to read the epochs' information immediately after the marker.

Altogether, as the latency of the overall system (the headset, wireless communication, and software) is unknown, and more importantly, delay in human response is inevitable, this question was raised: *Where it is best time to start incorporating samples within each epoch?* To answer this question, the proposed framework seeks for the best start-time and the best stop-time within each epoch, and readjusts the epochs by trimming the segments before the former and after the latter. In this fashion, the non-informative parts of the epochs are discarded, therefore, no (or less) irrelative

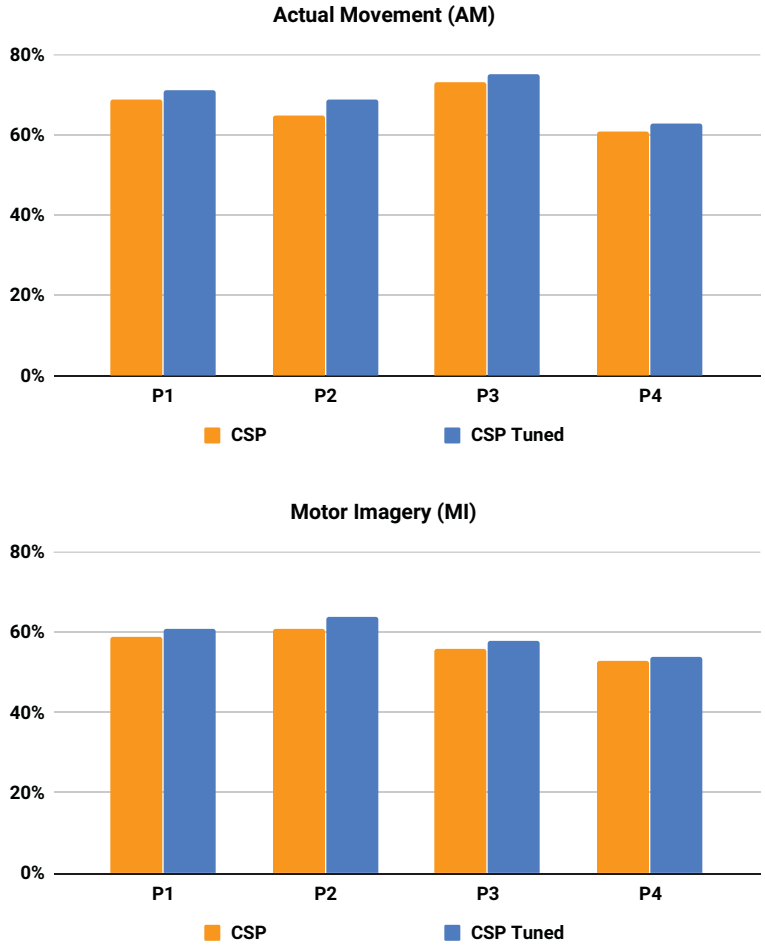

Figure 4.1: *Protocols P1 to P4* for *Actual Movement (AM)* and *Motor Imagery (MI)*: Average classification accuracy for 10 subjects: Comparison between *Tuned CSP* and *Conventional CSP*.

time samples are taken into account during the processing module of a BCI.

4.3.1 Trimming Framework Outline and Simulation

To evaluate the proposed framework, the raw datasets of BCI competition III-IVa were used. These datasets consist of data collected from 5 healthy subjects, (aa, al, av, aw, ay) and the EEG headset utilized had 118 channels with sampling frequency of 1000Hz. Subjects are asked to perform (L) left hand, (R) right hand as responses to visual cues indicated for 3.5 seconds, followed by 1.5 second rest, for a total of 280 trials. These three steps, without applying any dimensionality reduction technique, were deliberately adopted for the sole purpose of enabling a better exhibition of the effects of the trimming step, which comes before feature extraction.

- **Preprocessing**: Raw dataset of BCI Competition III-IVa were filtered via a fifth order*Butterworth* filter, to remove the DC gain and to pick the information within 7-30Hz. After that, the epochs were arranged and then smoothed using the weighted moving average method by a window size of 10-time samples. Then, the dataset is downsampled to keep one sample out of each batch of 10. Finally, within each epoch, there are 350-time samples.

- **Feature Extraction**: The CSP method is used to extract features. Two eigenvectors are taken for the purpose of construction of the whitening matrix. Also, in respect to the rule of thumb mentioned in Chapter 2, both scenarios of splitting the dataset into training and test trials are implemented: once with 60% of the trials for training, which is 168 trials out of 280, and the remaining 112 trials were tested by the trained classifier. Another run, 196 trials (70% of the trials) were put aside for training the classifier, and that leaves 84 trials in the test dataset.

- **Classification**: Both LDA and QDA models with 5-fold cross validation were utilized to classify the test sets.

As the main goal is to investigate whether or not a trimmed epoch would contribute to better classification accuracy, the following three different scenarios are considered:

Scenario 1: The trimming process starts at the beginning of each epoch. In other words, the trimming method considers a fixed time interval as the duration of the target user's response, and this interval slides over the epoch from the first time sample up to the point where the last time sample of the fixed interval is placed over the epoch's last time sample. This scenario assumes that the BCI user does not respond to the stimuli right away, and there is a delay in his/her reaction. Therefore, *Scenario 1* seeks to find the time sample which reflects the beginning of the user's response, via evaluation in each iteration of sliding the time interval. The algorithm is elaborated in Algorithm 4.

Scenario 2: In this scenario, the epochs are trimmed from the ending time samples. The trimming process begins from the last time sample of the epoch, and while considering a fixed time interval for the user's response, the interval slides backward, until the first time sample of the interval is placed over the first time sample of the epoch. The motivation behind this scenario is the fact that some subjects do begin their reaction to the stimuli right after noticing it. However, they stop responding

Algorithm 4 TRIMMING LOOP - SCENARIO 1

Input: The matrices of half of the training trials as $\boldsymbol{X}_{Ntrain/2}$.
For T_start = 1:5 (or 0.05s):200 (or 2s)

- T_stop = T_start + 150 (or 1.5s);
- $\boldsymbol{X}_{Ntrain/2_i}$ = $\boldsymbol{X}_{Ntrain/2_i}$(:,T_start:T_stop,:);
- The CSP is applied;
- The classification is carried out with LDA.

End

- The accuracies of classification with different T_start are evaluated: Best_T_start is extracted.
- The N_t for all epochs is readjusted based on the previous step.

soon after one or two repetitions of the MI task in their mind. Similar to *Scenario 1*, *Scenario 2* strives to find the time sample which projects that the user has stopped reacting to the stimulus. The algorithm is elaborated in Algorithm 5.

Scenario 3: In this scenario, a degree of flexibility is added to the trimming process by merging the previous two scenarios. In *Scenario 3*, the trimming process is flexible in terms of the time interval, and the start time sample and the stop time sample are both found based on a grid search over all epoch time samples. This design roots in the fact that some subjects might not have a consistent behavior while responding to the stimuli, therefore, the entire epoch has to be assessed and the time interval which includes the most information regarding the subject's MI intention will be extracted. The algorithm is elaborated in Algorithm 6.

For the implementation of the aforementioned three scenarios, we considered two rounds of processing, one with 60% of the datasets as the training dataset and the rest was put aside for the testing purpose. The other round, with 70% of the trials used for training the classification model and the remainder of the data tested by the learned model. For both rounds, the trimming loop was applied to half of the training trials, essentially, every other trial in the training dataset was taken into account as the input of this loop. For all three scenarios, we assumed that the minimum time interval required to respond to the stimuli and perform the MI task is 1.5 seconds, or in this specific case, 150-time samples of the preprocessed epoch data. As a final note, these three steps are implemented without deliberately applying any dimensionality reduction technique or introducing complexity into the classical algorithms. The rationale behind the adopted approach is our sole purpose of enabling a better exhibition of the effects of the trimming step, which comes before the feature extraction.

Algorithm 5 TRIMMING LOOP - SCENARIO 2

Input: The matrices of half of the training trials as $\boldsymbol{X}_{Ntrain/2}$.
For T_stop = 350:5 (or 0.05s):200 (or 2s)
- T_start = T_stop - 150 (or 1.5s);
- $\boldsymbol{X}_{Ntrain/2_i}$ = $\boldsymbol{X}_{Ntrain/2_i}$(:,T_start:T_stop,:);
- The CSP is applied;
- The classification is carried out with LDA.

End
- The accuracies of classification with different T_start are evaluated: Best_T_start is extracted.
- The N_t for all epochs is readjusted based on the previous step.

Algorithm 6 TRIMMING LOOP - SCENARIO 3

Input: The matrices of half of the training trials as $\boldsymbol{X}_{Ntrain/2}$.
For T_stop = 350:5 (or 0.05s):200 (or 2s)
- $\hat{\boldsymbol{X}}_{Ntrain/2_i}$ = $\boldsymbol{X}_{Ntrain/2_i}$(:,T_start:end:);
- The CSP is applied;
- The classification is carried out with LDA.

End
- The accuracies of classification with different T_start are evaluated: Best_T_start is extracted.
- The N_t for all epochs is readjusted based on the previous step.

While *"The trimming edges are within the epoch size"*
- T_Stop = Best_TStart + LoopCount $\times$ 5 (or 0.05s);
- $\hat{\hat{\boldsymbol{X}}}_{Ntrain/2_i}$ = $\hat{\boldsymbol{X}}_{Ntrain/2_i}$(:,1:T_stop,:);
- The CSP is applied;
- The classification is carried out with LDA.

End
- The accuracies of classification with different T_stop are evaluated: Best_T_stop is extracted.
- The N_t for all epochs is readjusted based on the previous step.

4.3.2 Simulation Results

Tables 4.2-4.4 show the accuracies of classification using both the proposed and the conventional framework through all three scenarios. As the results exhibit, the trimming loop significantly improves the performance of the kernel, and the average accuracy has been increased notably.

As can be seen in the results of Scenario 1, the trimming loop extracted the best T_start and, correspondingly, the best T_stop is extracted, adaptively for each subject. The time samples found are almost the same as those found via Scenario 2, which shows that the two of them succeeded in retrieving the time interval within which the concentration of the subject on performing the MI task is maximum. Both of these scenarios successfully enhance the classification accuracy for all subjects, which was the main purpose of the proposed framework. Regarding Scenarios 1 and 2,

Subjects TrainSet 60% TestSet 40%	Method	CSP LDA Classifier	CSP QDA Classifier	CSP SVM Classifier	Best_TStart (TimeSample)	Best_TStop (TimeSample)
AY	CSP + Trim	95% ± 5	94% ± 6	89% ± 7	201	350
	Conventional CSP	57% ± 9	60% ± 6	51% ± 2		
AW	CSP + Trim	94% ± 4	93% ± 3	94% ± 3	71	220
	Conventional CSP	46% ± 7	48% ± 2	48% ± 2		
AV	CSP + Trim	53% ± 8	50% ± 6	53% ± 2	1	150
	Conventional CSP	51% ± 4	51% ± 2	53% ± 9		
AL	CSP + Trim	95% ± 2	96% ± 5	95% ± 8	66	215
	Conventional CSP	92% ± 5	90% ± 4	92% ± 6		
AA	CSP + Trim	70% ± 6	70% ± 5	70% ± 7	161	310
	Conventional CSP	57% ± 8	51% ± 8	56.% ± 7		
Average	**CSP + Trim**	**81%**	**81%**	**80%**	**100**	**249**
	Conventional CSP	**61%**	**60%**	**60%**		

Subjects TrainSet 70% TestSet 30%	Method	CSP LDA Classifier	CSP QDA Classifier	CSP SVM Classifier	Best_TStart (TimeSample)	Best_TStop (TimeSample)
AY	CSP + Trim	90% ± 4	90% ± 4	86% ± 8	61	210
	Conventional CSP	56.% ± 3	59% ± 6	56.% ± 2		
AW	CSP + Trim	84% ± 3	86% ± 4	86% ± 4	46	195
	Conventional CSP	81% ± 3	84% ± 5	82% ± 6		
AV	CSP + Trim	50% ± 6	48% ± 7	49% ± 6	6	155
	Conventional CSP	48% ± 7	49% ± 8	48% ± 8		
AL	CSP + Trim	98% ± 4	100% ± 3	99% ± 3	106	255
	Conventional CSP	90% ± 3	90% ± 7	90% ± 4		
AA	CSP + Trim	68% ± 4	68% ± 5	68% ± 6	141	290
	Conventional CSP	63% ± 4	65% ± 8	61% ± 5		
Average	**CSP + Trim**	**78%**	**78%**	**78%**	**72**	**221**
	Conventional CSP	**68%**	**69%**	**67%**		

Figure 4.2: Performance comparison of the proposed trimming framework trained with either 60% or 70% of training trial based on Scenario 1.

our proposed method shows that the interval within each subject is the same and the similar results are obtained.

Finally, in the Scenario 3, after determining the best T_start, the best T_stop is calculated, in other words, we do not fix the duration and left this task to the method. However, although Scenario 3 is successful in enhancing the overall classification accuracies, it fails to outperforming the first two scenarios. Based on the results in Scenario 3, the most strong reaction of the subject can be found by the "for" loop, and the remainder of the epoch is trimmed. It is clear that this scenario is not as effective as the previous scenarios (1 and 2). Therefore, in the case of adding the trimming step to their framework with scenario 3, the BCI developers should fix the minimum time interval

Subjects TrainSet 60% TestSet 40%	Method	CSP LDA Classifier	CSP QDA Classifier	CSP SVM Classifier	Best_TStart (TimeSample)	Best_TStop (TimeSample)
AY	CSP + Trim	95% ± 5	94% ± 6	89% ± 7	191	340
	Conventional CSP	57% ± 9	60% ± 7	51% ± 2		
AW	CSP + Trim	87% ± 2	87% ± 2	87% ± 3	66	215
	Conventional CSP	46% ± 6	48% ± 2	48% ± 2		
AV	CSP + Trim	50% ± 7	49% ± 5	49% ± 5	6	155
	Conventional CSP	51% ± 4	51% ± 2	53% ± 9		
AL	CSP + Trim	95% ± 6	96% ± 5	95% ± 2	66	215
	Conventional CSP	92% ± 5	90% ± 4	92% ± 6		
AA	CSP + Trim	76% ± 5	76% ± 4	75% ± 7	106	255
	Conventional CSP	57% ± 8	51% ± 2	56% ± 2		
Average	**CSP + Trim**	**81%**	**80%**	**79%**	**87**	**236**
	Conventional CSP	**61%**	**60%**	**60%**		

Subjects TrainSet 70% TestSet 30%	Method	CSP LDA Classifier	CSP QDA Classifier	CSP SVM Classifier	Best_TStart (TimeSample)	Best_TStop (TimeSample)
AY	CSP + Trim	89% ± 5	86% ± 6	86% ± 5	176	325
	Conventional CSP	56% ± 3	59% ± 6	56% ± 2		
AW	CSP + Trim	86% ± 2	86% ± 3	88% ± 5	61	210
	Conventional CSP	81% ± 3	84% ± 5	82% ± 6		
AV	CSP + Trim	50% ± 7	45% ± 9	46% ± 2	1	150
	Conventional CSP	48% ± 7	49% ± 8	48% ± 8		
AL	CSP + Trim	98% ± 5	100% ± 3	99% ± 3	106	255
	Conventional CSP	90% ± 4	90% ± 7	90% ± 4		
AA	CSP + Trim	64% ± 4	64% ± 5	67% ± 2	41	190
	Conventional CSP	63% ± 6	65% ± 8	61% ± 5		
Average	**CSP + Trim**	**77%**	**76%**	**77%**	**77**	**226**
	Conventional CSP	**68%**	**69%**	**67%**		

Figure 4.3: Performance comparison of the proposed trimming framework trained with either 60% or 70% of training trial based on Scenario 2.

based on the design of their system and the feedback received from pilot studies.

4.4 Adaptive Subjected-Specific Bayesian Spectral Filtering for Signal Trial EEG Classification

It is well known in data science and machine learning that feature extraction step is one of the most crucial steps for the development of an effective algorithm. More importantly and in the context of rehabilitation, any BCI system processing pipeline will benefit from better classification accuracies, if the feature extraction step is done in a systematic, optimized manner. We propose a

Subjects TrainSet 60% TestSet 40%	Method	CSP LDA Classifier	CSP QDA Classifier	CSP SVM Classifier	Best_TStart (TimeSample)	Best_TStop (TimeSample)
AY	CSP + Trim	69% ± 9	63% ± 9	67% ± 2	191	196
	Conventional CSP	57% ± 9	60% ± 7	51% ± 2		
AW	CSP + Trim	53% ± 3	53% ± 2	53% ± 4	116	121
	Conventional CSP	46% ± 7	48% ± 2	48% ± 2		
AV	CSP + Trim	54% ± 3	56% ± 3	52% ± 3	1	6
	Conventional CSP	51% ± 5	51% ± 2	53% ± 9		
AL	CSP + Trim	94% ± 3	94% ± 3	94% ± 3	56	181
	Conventional CSP	92% ± 5	90% ± 4	92% ± 6		
AA	CSP + Trim	61% ± 8	59% ± 2	60% ± 2	96	106
	Conventional CSP	57% ± 8	51% ± 2	56% ± 2		
Average	**CSP + Trim**	**66%**	**65%**	**65%**	**73**	**122**
	Conventional CSP	**61%**	**60%**	**60%**		

Subjects TrainSet 70% TestSet 30%	Method	CSP LDA Classifier	CSP QDA Classifier	CSP SVM Classifier	Best_TStart (TimeSample)	Best_TStop (TimeSample)
AY	CSP + Trim	83% ± 5	81% ± 3	81% ± 3	166	176
	Conventional CSP	56% ± 4	59% ± 6	56% ± 2		
AW	CSP + Trim	45% ± 4	50% ± 5	51% ± 2	51	56
	Conventional CSP	81% ± 3	84% ± 6	82% ± 6		
AV	CSP + Trim	52% ± 6	55% ± 5	52% ± 7	11	21
	Conventional CSP	48% ± 7	49% ± 8	48% ± 8		
AL	CSP + Trim	98% ± 4	99% ± 3	98% ± 3	121	226
	Conventional CSP	90% ± 4	90% ± 7	90% ± 4		
AA	CSP + Trim	62% ± 6	62% ± 6	62% ± 8	61	66
	Conventional CSP	63% ± 6	65% ± 8	61% ± 5		
Average	**CSP + Trim**	**68%**	**69%**	**69%**	**82**	**109**
	Conventional CSP	**68%**	**69%**	**67%**		

Figure 4.4: Performance comparison of the proposed trimming framework trained with either 60% or 70% of training trial based on Scenario 3.

subject-specific filtering framework, referred to as the regularized double-band Bayesian (R-B2B) spectral filtering. The R-B2B couples three main feature extraction categories, namely filter-bank solutions, regularized techniques, and optimized Bayesian mechanisms to enhance the classification accuracy by simultaneously taking advantage of the three processing techniques in motor imagery (MI) and actual movement tasks in EEG employed for Brain-Computer Interfaces (BCIs). Furthermore, real data collection experiments are performed to investigate different effects of stimulus on the overall performance of the proposed R-B2B. In this regard, four different protocols are designed and implemented by introducing visual and voice stimuli. Finally, the work investigates effects of

adaptive trimming of EEG epochs resulting in an adaptive and subject-specific solution. Experimental results show that the proposed R-B2B filter noticeably outperforms its counterparts. The overall procedure in the R-B2B framework is divided into two parts: Regularized Common Spatial Pattern (RCSP) and Double-band Bayesian Spatio-Spectral Filter Optimization as described below in detail.

4.4.1 EEG Data Collection and Pre-Processing

We consider supervised learning from EEG signals based on the available set of EEG epochs (trials) denoted by $\boldsymbol{X}_i \in \mathbb{R}^{N_{\text{ch}} \times N_{\text{t}}}$, for $(1 \leq i \leq N_{\text{Trial}})$, where N_{Trial} is the total number of epochs used for processing; N_{ch} is the number of EEG channels (electrodes), and; N_{t} is the number of time samples collected from each electrode in each trial. The training dataset is denoted by $\{(\boldsymbol{X}_i, L_i)\}$, for $(1 \leq i \leq N_{\text{Trial}})$, where L_i represents the label corresponding to the ith trial. The EEG dataset used in this research is collected from 10 healthy subjects performing actual movement and motor imagery (MI) tasks. Depending on the presented cue, the subjects are asked to perform either a left-hand movement, referred to as Class C1, or a right-hand movement, referred to as Class C2, depending on a visually presented cue. A portable and wireless biosignal acquisition system, g.Nautilus from g.tech Medical Engineering g.Tech: g.Nautilus [2018] is used to collect brain signals. The g.tech system consists of a cap with 32 bipolar active wet electrodes that are uniformly distributed according to the international 10-20 standard. For each protocol $P1$, $P2$, $P3$, and $P4$ described in Chapter 3, each subject was asked to perform actual movements of the left or right hand or imagine doing these movements. The total duration of the reaction time for each trial is 3 seconds.

During the pre-processing step, Raw EEG signals are filtered by applying a *notch* filter at 60Hz to remove the power line interference, followed by the *Chebyshev-II* filter of order 10 to collect information within 0.5 to 100 Hz. Finally, the signal is filtered using the*Chebyshev-II* filter of order 20 within 4 to 40 Hz.

4.4.2 Regularized-Double band Spatio-Spatial Filtering

Feature extraction is performed after the preprocessing step and aims at characterizing the acquired data to help identify its correspondence to one of several predefined mental activities. For performing subject-specific feature extraction, we extend the B2B framework discussed in Section 2.2.6 by using a regularized version of the CSP technique instead of the conventional one used in the original work Shahtalebi and Mohammadi [2018]. The main differences between the proposed method in comparison to previous works are as follows:

- **I:** Instead of a single spectral filtering in the Bayesian Spatio-spectral Filter Optimization Framework Suk and Lee [2013], characteristics of a filter bank (2 frequency bands) are optimized, which is mainly stored in μ band (8-13Hz) and β band (13-30Hz).
- **II:** Taking advantage of the benefits of filter bank methods and the benefits of optimized spectral filtering methods, in a way to reach higher accuracies.
- **III:** Using Regularized version of CSP instead of simple CSP to address the problem of small training dataset by reducing both bias and variance using regularized parameters.

In the proposed R-B2B approach, the spectral filters are modelled with *a priori* probability denoted by $p(\boldsymbol{\mathcal{B}}^{\mathcal{D}})$ over the reference random variable $\boldsymbol{\mathcal{B}}^{\mathcal{D}}$. The cutoff frequencies of a bandpass filter is defined as $\boldsymbol{\mathcal{B}}^{\mathcal{D}} \triangleq [b_s, b_m, b_e]$. then the output $\boldsymbol{\mathcal{Z}}_l$ of each bandpass filter is modeled by the convolution of the input signal $\boldsymbol{X}_i$ for $(1 \leq i \leq N_{\text{Trial}})$ with system $h_l(k)$, $l \in \{1, 2\}$, and is given by

$$\boldsymbol{\mathcal{Z}}_{i,l}(k) = h_l(k) \circledast \boldsymbol{X}_i, \quad l \in \{1, 2\}. \tag{33}$$

The posterior distribution over $\boldsymbol{\mathcal{B}}^{\mathcal{D}}$ is denoted by $p(\boldsymbol{\mathcal{B}}^{\mathcal{D}}|\boldsymbol{X}_i, \Omega_i)$ and is constructed from single-trial EEG recording $\boldsymbol{X}_i$ for $(1 \leq i \leq N_{\text{Trial}})$, The particle-based approximation of the posterior distribution is computed as follows

$$p(\boldsymbol{\mathcal{B}}^{\mathcal{D}}(k)|\boldsymbol{X}_i, \Omega_i) = \frac{p(\boldsymbol{X}_i, \Omega_i|\boldsymbol{\mathcal{B}}^{\mathcal{D}}(k))p(\boldsymbol{\mathcal{B}}^{\mathcal{D}}(k))}{p(\boldsymbol{X}_i, \Omega_i)}. \tag{34}$$

Based on Eq. (33), the likelihood and evidence are, respectively, defined as $p(\boldsymbol{\mathcal{Z}}_{i,l}(k), \Omega_i|\boldsymbol{\mathcal{B}}^{\mathcal{D}}(k))$ and $p(\boldsymbol{\mathcal{Z}}_{i,l}(k), \Omega_i)$ by replacing the raw EEG signal $\boldsymbol{X}_i$ with its bandpass-filtered version $\boldsymbol{\mathcal{Z}}_{i,l}(k)$ as

$$p(\boldsymbol{\mathcal{B}}^{\mathcal{D}}(k)|\boldsymbol{\mathcal{Z}}_{i,l}(k), \Omega_i) = \frac{p(\boldsymbol{\mathcal{Z}}_{i,l}(k), \Omega_i|\boldsymbol{\mathcal{B}}^{\mathcal{D}}(k))p(\boldsymbol{\mathcal{B}}^{\mathcal{D}}(k))}{p(\boldsymbol{\mathcal{Z}}_{i,l}(k), \Omega_i)}, \tag{35}$$

where $\boldsymbol{\mathcal{B}}^{\mathcal{D}}(k) = \{b_s(k), b_m(k), b_e(k), \pi(k)\}$, $(1 \leq k \leq N_p)$, denotes particle set k and contains particles weight $\pi(k)$. It represents a single filterbank among the N_p instances (particles) used. The spectral filtering step is followed by computing regularized CSP features. In regularized CSP, the covariance matrix for each trial $\boldsymbol{X}_i$, $(1 \leq i \leq N_{\text{Trial}})$ is used for discrimination of two classes and is computed as $\boldsymbol{\Sigma}_i = (\boldsymbol{X}_i\boldsymbol{X}_i^T)/(\text{Tr}(\boldsymbol{X}_i\boldsymbol{X}_i^T))$. The generic average spatial covariance matrix for each class is then calculated from

$$\boldsymbol{\Sigma}^{(c)} = \frac{1}{N_{\text{Trial}}^{(c)}} \sum_{i=1}^{N_{\text{Trial}}^{(c)}} \boldsymbol{\Sigma}_i^{(c)}, \qquad c \in \{0, 1\}, \tag{36}$$

where $N_{\text{Trial}}^{(c)}$ is the number of epochs/trials belonging to class $c \in \{0, 1\}$. We also compute a subject-specific average covariance matrix denoted by $\boldsymbol{\Sigma}^{(c,j)}$ for subject j, $(1 \leq j \leq N_{\text{Sub}})$, with N_{Sub} denoting the total number of available subjects. Note that, $\boldsymbol{\Sigma}^{(c,j)}$ is computed from Eq. (36), where the summation is performed over the number of epochs available for Subject j. The regularized subject-specific covariance matrix is then computed as follows

$$\hat{\boldsymbol{\Sigma}}^{(c,j)}(\beta, \gamma) = (1 - \gamma)\hat{\boldsymbol{\Sigma}}^{(c,j)}(\beta) + \frac{\gamma}{N_{\text{ch}}} \text{tr}\big[\hat{\boldsymbol{\Sigma}}^{(c,j)}(\beta)\big]\boldsymbol{I} \tag{37}$$

where β and γ, are regularization parameters, $\boldsymbol{I}$ is an $(N_{\text{ch}} \times N_{\text{ch}})$ identity matrix, and $\hat{\boldsymbol{\Sigma}}^{(c,j)}(\beta)$ is defined as

$$\hat{\boldsymbol{\Sigma}}^{(c,j)}(\beta) = \frac{(1 - \beta)\boldsymbol{\Sigma}^{(c,j)} + \beta\boldsymbol{\Sigma}^{(c)}}{(1 - \beta)N_{\text{Trial}}^{(c,j)} + \beta N_{\text{Trial}}^{(c)}}, \qquad c \in \{0, 1\}. \tag{38}$$

For Subject j, the composite spatial covariance matrix is computed as

$$\bar{\boldsymbol{\Sigma}}^{(j)}(\beta, \gamma) = \hat{\boldsymbol{\Sigma}}^{(0,j)}(\beta, \gamma) + \hat{\boldsymbol{\Sigma}}^{(1,j)}(\beta, \gamma). \tag{39}$$

The next step is to perform an eigenvalue decomposition on the regularized composite spatial covariance matrix for subject j, ($1 \leq j \leq N_{\text{Sub}}$). Representing $\bar{\boldsymbol{\Sigma}}^{(j)}(\beta,\gamma) = \boldsymbol{V}^{(j)}\boldsymbol{\lambda}^{(j)}[\boldsymbol{V}^{(j)}]^T$, where $\boldsymbol{V}^{(j)}$ is the matrix containing eigenvectors associated with the composite covariance and $\boldsymbol{\lambda}^{(j)}$ is the diagonal matrix of its corresponding eigenvalues. Based on Ang et al. [2012] and using Matlab notation, the spatial filter is computed as $\boldsymbol{W}^{(j)} = \text{eig}(\bar{\boldsymbol{\Sigma}}^{(j)}(\beta,\gamma), \hat{\boldsymbol{\Sigma}}^{(1,j)}(\beta,\gamma))$. The spatial filter $\boldsymbol{W}_i^{(j)}$ is then used to form the decomposition of each trial $\boldsymbol{X}_i$, ($1 \leq i \leq N_{\text{Trial}}$). The spatial decomposition for the k'th particle, ($1 \leq k \leq N_p$), is given by $\boldsymbol{\mathcal{W}}_{i,l}^{(j)}(k) = \left[\boldsymbol{W}_l^{(j)}(k)\right]^T \boldsymbol{Z}_{i,l}^{(j)}(k)$, where the first and last l rows of matrix $\boldsymbol{W}_i^{(j)}(k)$ which is associated with the largest eigenvalues that maximizes the difference of variance between the two classes. Finally, the feature extraction is given by

$$\boldsymbol{f}_{i,l}^{(j)}(k) = \log\left(\frac{\text{var}(\boldsymbol{\mathcal{W}}_{i,l}^{(j)})}{\sum_i \text{var}(\boldsymbol{\mathcal{W}}_{i,l}^{(j)})}\right), \tag{40}$$

and the features of the two parallel pipelines are merged into one single feature vector as

$$\boldsymbol{f}_i^{(j)}(k) = \left[f_{i,l}^{(j)}(k)|_{l=1}, \boldsymbol{f}_{i,l}^{(j)}(k)|_{l=2}\right]. \tag{41}$$

This completes description of the R-B2B framework.

4.4.3 Experimental Results

We performed several experiments on the same subjects and compared four different protocols of stimulations based on visual and voice stimulus. For visual stimulus, left or right hand icons are shown to the subject corresponding to one of the two classes of suggested movement (tasks). For voice stimulus, the two tasks are differentiated with a voice saying left or right. Based on the proposed method, tuned parameters and comparing different protocols, the experimental results clearly show that:

Proposed R-B2B Framework: The performance of the proposed R-B2B framework is compared with B2B in Table 4.2. In training and evaluation sessions used in both frameworks, segments of 3 seconds (corresponding to 1500 time-samples) from each trial are selected as input. After spatial

filtering, the first and last three rows of the signals (6 eigenvectors) are used for feature extraction. In the classification step, a linear SVM classier is used by investigating the optimized band-limits for the spectral filters in the R-B2B framework. As shown in Fig. 4.2, the proposed R-B2B framework achieves promising results in comparison to the B2B.

Tuned Parameters Affection: A set of best parameters, i.e., *Tuned TimeSample* related to each subject, selecting *6 number of Eigen Features*, and *SVM Classifier*, is obtained from the grid-search method. In comparison to the conventional CSP, the result could yield relatively the best performance for both tasks. Moreover, we realized that in Protocol P1 (where the reaction and visual stimulations happened simultaneously), much higher performance is observed at the beginning of recording (less than 250 TimeSamples) and the performance gain decreases at the end of the recording. We can thus infer that the effect of visual stimulation gives us an "unreliable" performance gain that does not represent motor imagery or actual movement tasks.

Protocol's Comparison: For Protocol P2, we reduce the impact of visual stimulation by creating a 1.25-second gap between the pop-up hands representing the visual stimulus and subject's response. This gap which led to having the most significant performance for MI tasks. Moreover, during Protocols P3 and P4, we asked subjects to close their eyes and react appropriately as they listen to the voice stimulus. In Protocol P3, voice stimulus is separated from the recording time and in Protocol P4, the visual stimulation is maintained during the recording. The results for those protocol P3 and P4 are shown that the effect of voice stimulation for Protocol P3 gave us the significant performance in actual movement and MI compared with protocol P4.

In summary, four different protocols, followed by different feature extraction approaches, are compared in this section for MI and actual movement tasks. The dataset created for our experiments is used to evaluate the performance of the proposed framework. Our results clearly show better performance for protocol P1 in both tasks; however, this might not be reliable. Protocols P2 and P3 (based on different tasks), which are isolating brain responses and the subsequent EEG recording from induced stimulation (whether speech or visual) exhibit superior performance. Moreover, P3 shows better performance in performing actual movements, while P2's results are more powerful in performing MI tasks.

Table 4.2: *Protocols P1 to P4* for *Actual Movement (AM)* and *Motor Imagery (MI)*: Average classification performance for 10 subjects: comparison between the proposed R-B2B and the original B2B techniques.

Avg. Accuracy of 10 Subjects	Tasks	B2B-CSP	B2B-RCSP
Protocol 1	AM	72%	82%
Protocol 2	AM	70%	74%
Protocol 3	AM	77%	80%
Protocol 4	AM	67%	77%
Protocol 1	MI	63%	72%
Protocol 2	MI	67%	69%
Protocol 3	MI	59%	61%
Protocol 4	MI	56%	58%

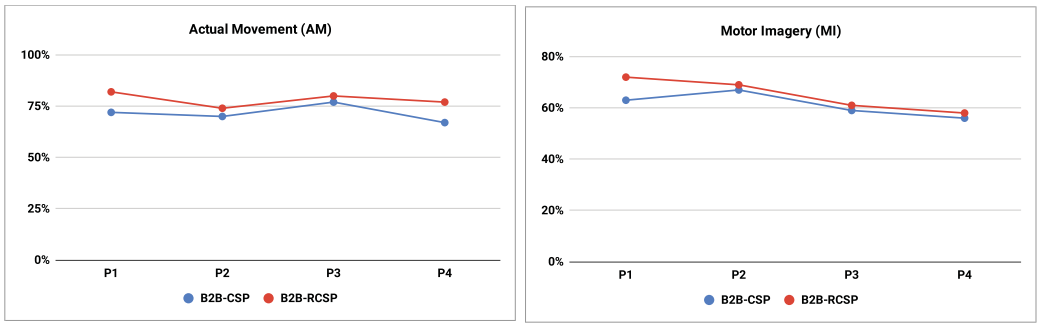

Figure 4.5: *Protocols P1 to P4* for *Actual Movement (AM)* and *Motor Imagery (MI)*: Average classification performance for 10 subjects: comparison between the proposed R-B2B and the original B2B techniques

4.5 Conclusion

In this chapter, first, an algorithm is proposed aiming to investigate whether or not a trimmed epoch would contribute to better classification accuracy. The experimental results show that under all three scenarios, the proposed framework is significantly effective and contributes to a much better performance in comparison to the conventional framework. A subject-specific filtering framework, referred to as the regularized double-band Bayesian (R-B2B) spectral filtering, is then proposed, which couples three main feature extraction categories (namely filter-bank solutions, regularized

techniques, and optimized Bayesian mechanisms) to enhance the classification accuracy. Our experiments indicate that the R-B2B approach provides notable performance gain over the B2B framework in binary BCI classification problems. Finally, future work shall investigate more complicated feature extraction and classification methods, and additionally, the dimensionality reduction methods will be added to evaluate the performance of the algorithm.

Chapter 5

Conclusion and Future Research work

The chapter concludes the thesis with a list of significant contributions made in the dissertation and proposed directions for future work.

5.1 Summary of Contributions

In what follows, I briefly outlined the main contributions of the thesis:

(1) Biomedical Signal Acquisition with Electroencephalogram (EEG): A BCI system is implemented and a dataset is collected from 10 healthy volunteer subjects performing an actual movement (AM) and Motor Imagery (MI) tasks. The constructed dataset is used to study different aspects of the proposed subject-specific filtering frameworks. To investigate the effects of different stimulus on the overall achievable performance, four different protocols based on visual and voice stimulus are introduced. The brain activity were recorded by 32 bipolar recordings with a sampling frequency rate of 500Hz. Each trial is started with the presentation of a fixation cross in the center of the screen, followed by four different paradigms and randomize breaks between consecutive trials (ranging between 2 to 3 seconds). In three experimental runs of 100 trials, each subject was asked to perform actual movements of the left or right hand or imagine about doing these movements. In terms of evaluating the collected datasets, first EEG signals are filtered by applying a notch filter at 60Hz to remove the power line interference, followed by the Chebyshev-II filter of order 10 to collect information within 0.5 to 100 Hz. Finally, the pre-processed signals are filtered using the

Chebyshev- II filter of order 20 within 4 to 40 Hz. Thereafter, the signals are downsampled to keep one sample out of a batch of 10 samples. Common Spatial Patterns (CSP) is employed as the feature extraction procedure to extract spatial features of ERD/ERS coupled with a trimming mechanism to choose a set of optimal hyper-parameters.

The experimental results based on different protocols show that by reducing the impact of visual stimulation, i.e., creating a 1.25 second gap between the pop-up hands representing the visual stimulus and subject's response (Protocol P2 in the study) more significant performance for MI tasks is gained. In Protocol P3, voice stimulus is separated from the recording time and in Protocol P4, the visual stimulus is maintained during the recording. The results for Protocols P3 and P4 illustrate that voice stimulation provides significant performance improvement while actual movement and motor imagery are performed in comparison to Protocol P4. Among the different suggested protocols, it is observed that the best performance in actual movement task comes from protocol P3 (voice popped up by ignoring the voice stimulation including a 1.25 second gap between hearing the voice stimulus and response recording). However, the best performance in motor imagery is achieved in Protocol P2 (hands popped up by ignoring the visual stimulation including a 1.25 second gap between visual stimulus and response recording).

(2) Improving Accuracy of EEG-based BCIs via Tuning Mechanism: Intuitively speaking, this framework significantly effects the resulting model's performance. A tuning mechanism is proposed to identify a set of optimal parameters, in terms of the overall achievable performance during the validation step. In particular, for each subject, the tuning mechanism first obtains a time window within each EEG epoch (in terms of the start and end sample of the interval). Intuitively speaking, the extracted interval is associated with the most informative section of epochs associated with a particular subject. In comparison to conventional techniques, application of tuning mechanism provides relatively high performance improvements for performing both AM an MI tasks. Moreover, we realized that in Protocol P1 (where the reaction and visual stimulations happened simultaneously), much higher performance is observed at the beginning of the recordings (less than 250 time samples) and the performance gain decreases at the end of recording. We can thus infer that the effect of visual stimulation gives us an "unreliable" performance gain that does not represent AM or MI.

(3) Improving the Accuracy of MI EEG-based BCIs Through Trimming the Epochs: This thesis proposed a readjustment method to trim the epochs such that most informative parts of the signals are extracted and the segments of the epochs which do not include the response of the subjects to the stimuli would be discarded. This approach was tested by adding the trim step to a conventional CSP-based framework, and the results successfully prove the positive effect of this technique.

(4) Adaptive Subjected-Specific Bayesian Spectral Filtering for Signal Trial EEG Classification: A new feature extraction approach referred to as Subject-Specific Bayesian Spectral Filtering (R-B2B) framework is proposed. The proposed R-B2B offers the following benefits: (i) Instead of a single spectral filtering in the Bayesian Spatio-spectral filter Optimization framework, characteristics of a filter bank (2 frequency bands) are optimized; (ii) Higher accuracies are achieved by taking advantage of the benefits of filter bank methods and the benefits of optimized spectral filtering methods, and; (iii) Regularized version of the CSP approach is utilized, which addresses the problem of small training dataset by reducing both bias and variance of the model using regularized parameters. Experiments indicate that the R-B2B approach provides notable performance gain over the B2B framework in binary BCI classification problems.

5.2 Future Work

In the following, we sum up some of the potential directions for future research work:

- Considering the fact that through a cognitive process, the subject's brain learns to decrease the motor concentration while doing the same task repeatedly, one direction for future research is to consider possibility of using both visual and voice stimulus for performing motor imagery BCI experiments.

- In the processing pipeline developed through this thesis research work, I did not utilize artifact removal techniques aiming at removal of different sources of interference potentially encountered in the electroencephalogram (EEG) recordings including ocular, eye blinks, muscular, and cardiac artifacts. In this regards, one direction for future research is to incorporate artifact removal algorithms prior to applying the proposed solutions. For instance, EEG-Lab, which is a powerful Matlab graphics interface, can be used to remove different artifacts and further

improve the overall accuracy of developed approaches. Moreover, as it is easier to prevent artifacts from occurring to remove them from data, to have a more reliable dataset, experimenters should aim to record the dataset as artifact-free as possible. In this regards, one needs to control non-physiological artifacts by improving the protocol design/implementation such as choosing a comfortable subject chair, low light levels in the lab, a pleasant and engaged experimenter, and proper explanation of the recording protocol. Finally, some subjects may be drowsy, therefore, dividing experiments into runs of 10-15 minutes duration can be beneficial.

- Sample size computation is an essential aspect of neuroscience and clinical investigations. Increasing the number of subjects could result in having a more valid and reliable performance.
- Instead of using only a few number of Eigen features from the transforming matrix and ignoring the rest of the eigenvectors; the weighting approach can be used in terms of utilizing all CSP features to avoid information loss.
- More complicated classification models shall be tried out.
- The frameworks can be developed for multi-class BCIs.

References

Abo-Zahhad, M., Ahmed, S. M., & Abbas, S. N. (2015). A new eeg acquisition protocol for biometric identification using eye blinking signals. *International Journal of Intelligent Systems and Applications*, *7*(6), 48.

Aghaei, A. S., Mahanta, M. S., & Plataniotis, K. N. (2016). Separable common spatio-spectral patterns for motor imagery bci systems. *IEEE Transactions on Biomedical Engineering*, *63*(1), 15–29.

Ang, K. K., Chin, Z. Y., Wang, C., Guan, C., & Zhang, H. (2012). Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b. *Frontiers in neuroscience*, *6*, 39.

Angeli, C. A., Boakye, M., Morton, R. A., Vogt, J., Benton, K., Chen, Y., ... Harkema, S. J. (2018). Recovery of over-ground walking after chronic motor complete spinal cord injury. *New England Journal of Medicine*, *379*(13), 1244–1250.

Baxter, S., Enderby, P., Evans, P., & Judge, S. (2012). Barriers and facilitators to the use of high-technology augmentative and alternative communication devices: a systematic review and qualitative synthesis. *International Journal of Language & Communication Disorders*, *47*(2), 115–129.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., & Muller, K.-R. (2008). Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal processing magazine*, *25*(1), 41–56.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, *2*(2), 121–167.

Chen, Y.-Y., Lai, H.-Y., Lin, S.-H., Cho, C.-W., Chao, W.-H., Liao, C.-H., ... Lin, S.-Y. (2009). Design and fabrication of a polyimide-based microelectrode array: application in neural recording and repeatable electrolytic lesion in rat brain. *Journal of neuroscience methods*, *182*(1), 6–16.

Claesen, M., & De Moor, B. (2015). Hyperparameter search in machine learning. *arXiv preprint arXiv:1502.02127*.

Daly, J. J., & Wolpaw, J. R. (2008). Brain–computer interfaces in neurological rehabilitation. *The Lancet Neurology*, *7*(11), 1032–1043.

Decety, J. (1996). Do imagined and executed actions share the same neural substrate? *Cognitive brain research*, *3*(2), 87–93.

Edelman, B., Baxter, B., & B, H. (2016). Eeg source imaging enhances the decoding of complex right-hand motor imagery tasks. *IEEE Transactions on Biomedical Engineering*, *63*(1), 4–14.

electroencephalography book, H. (2018). Human electroencephalography book, chapter 10. examination of biolectrical signals accompanying brain function (eeg). *http://ttktamop.elte.hu/onlinetananyagok/ physiology practical/ch10s02.htmly*, 99–110.

Ferguson, K., Polando, G., Kobetic, R., Triolo, R. J., & Marsolais, E. (1999). Walking with a hybrid orthosis system. *Spinal cord*, *37*(11), 800.

Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American statistical association*, *84*(405), 165–175.

Garrett, D., Peterson, D. A., Anderson, C. W., & Thaut, M. H. (2003). Comparison of linear, nonlinear, and feature selection methods for eeg signal classification. *IEEE Transactions on neural systems and rehabilitation engineering*, *11*(2), 141–144.

Gill, M. L., Grahn, P. J., Calvert, J. S., Linde, M. B., Lavrov, I. A., Strommen, J. A., ... others (2018). Neuromodulation of lumbosacral spinal networks enables independent stepping after complete paraplegia. *Nature medicine*, *24*(11), 1677.

Grenander, U., Chow, Y.-s., & Keenan, D. M. (2012). *Hands: A pattern theoretic study of biological shapes* (Vol. 2). Springer Science & Business Media.

Grosse-Wentrup, M., Liefhold, C., Gramann, K., & Buss, M. (2009). Beamforming in noninvasive

brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, *56*(4), 1209–1219.

g.Tech: g.Nautilus. (2018). g.tech: g.nautilus wireless biosignal acquisition, http://www.gtec.at/products/hardware-and- accessories/g.nautilus-specs-feature.

Harrar, S. W., & Gupta, A. K. (2008). On matrix variate skew-normal distributions. *Statistics*, *42*(2), 179–194.

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia*, *33*(11), 1419–1432.

Khasnobish, A., Bhattacharyya, S., Konar, A., Tibarewala, D., & Nagar, A. K. (2011). A two-fold classification for composite decision about localized arm movement from eeg by svm and qda techniques. In *Neural networks (ijcnn), the 2011 international joint conference on* (pp. 1344–1351).

Kilgore, K., Peckham, P., & Keith, M. (1997). Th rope gb, wuolle ks, bryden am, hart ra.. an imphiucd upper-extremity neuroprosthesis. follow-up of patients. *Bone Joint Surg [Ami*, *79*, 533–5.

Kohavi, R., et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai* (Vol. 14, pp. 1137–1145).

Kübler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J. R., & Birbaumer, N. (2001). Brain–computer communication: Unlocking the locked in. *Psychological bulletin*, *127*(3), 358.

Kubota, M., Sakakihara, Y., Uchiyama, Y., Nara, A., Nagata, T., Nitta, H., ... Yanagisawa, M. (2000). New ocular movement detector system as a communication tool in ventilator-assisted werdnig-hoffmann disease. *Developmental Medicine & Child Neurology*, *42*(1), 61–64.

Kumar, S., Sharma, A., & Tsunoda, T. (2017). An improved discriminative filter bank selection approach for motor imagery eeg signal classification using mutual information. *BMC bioinformatics*, *18*(16), 545.

LaCourse, J. R., & Hludik, F. C. (1990). An eye movement communication-control system for the disabled. *IEEE Transactions on Biomedical Engineering*, *37*(12), 1215–1220.

Lemm, S., Müller, K.-R., & Curio, G. (2009). A generalized framework for quantifying the dynamics of eeg event-related desynchronization. *PLoS computational biology*, *5*(8), e1000453.

Lotte, F., & Guan, C. (2011). Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms. *IEEE Transactions on biomedical Engineering*, *58*(2), 355–362.

Lu, H., Eng, H.-L., Guan, C., Plataniotis, K. N., & Venetsanopoulos, A. N. (2010). Regularized common spatial pattern with aggregation for eeg classification in small-sample setting. *IEEE transactions on Biomedical Engineering*, *57*(12), 2936–2946.

Lu, H., Plataniotis, K. N., & Venetsanopoulos, A. N. (2009). Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition. *IEEE Transactions on Neural Networks*, *20*(1), 103–123.

Mason, S. G., & Birch, G. E. (2003). A general framework for brain-computer interface design. *IEEE transactions on neural systems and rehabilitation engineering*, *11*(1), 70–85.

Moghimi, S., Kushki, A., Marie Guerguerian, A., & Chau, T. (2013). A review of eeg-based brain-computer interfaces as access pathways for individuals with severe disabilities. *Assistive Technology*, *25*(2), 99–110.

Mohammadi, A., & Asif, A. (2013). Distributed particle filter implementation with intermittent/irregular consensus convergence. *IEEE Transactions on Signal Processing*, *61*(10), 2572–2587.

Moore, M. M. (2003). Real-world applications for brain-computer interface technology. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *11*(2), 162–165.

Müller-Gerking, J., Pfurtscheller, G., & Flyvbjerg, H. (1999). Designing optimal spatial filters for single-trial eeg classification in a movement task. *Clinical neurophysiology*, *110*(5), 787–798.

Nicolas-Alonso, L. F., & Gomez-Gil, J. (2012). Brain computer interfaces, a review. *sensors*, *12*(2), 1211–1279.

Nijholt, A., Tan, D., Pfurtscheller, G., Brunner, C., Millán, J. d. R., Allison, B., ... Müller, K.-R. (2008). Brain-computer interfacing for intelligent systems. *IEEE intelligent systems*, *23*(3), 72–79.

Park, S., Lee, D., & Lee, S. (2018). Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(2), 498–505.

Postelnicu, C., Talaba, D., & Toma, M. (2010). Brain computer interfaces for medical applications. *Bulletin of the Transilvania University of Brasov. Engineering Sciences. Series I*, *3*, 99.

Ramoser, H., Muller-Gerking, J., & Pfurtscheller, G. (2000). Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE transactions on rehabilitation engineering*, *8*(4), 441–446.

Raudys, S. J., & Jain, A. K. (1991). Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Transactions on Pattern Analysis & Machine Intelligence*(3), 252–264.

Reuderink, B., & Poel, M. (2008). Robustness of the common spatial patterns algorithm in the bci-pipeline. *University of Twente, Tech. Rep*.

Scherer, R., Muller, G., Neuper, C., Graimann, B., & Pfurtscheller, G. (2004). An asynchronously controlled eeg-based virtual keyboard: improvement of the spelling rate. *IEEE Transactions on Biomedical Engineering*, *51*(6), 979–984.

Shahtalebi, S., & Mohammadi, A. (2017a). Error correction output codding coupled with the csp for motor imagery bci systems. In *Signal processing conference (eusipco), 2017 25th european* (pp. 2071–2075).

Shahtalebi, S., & Mohammadi, A. (2017b). Ternary ecoc classifiers coupled with optimized spatio-spectral patterns for multiclass motor imagery classification. In *Systems, man, and cybernetics (smc), 2017 ieee international conference on* (pp. 2231–2236).

Shahtalebi, S., & Mohammadi, A. (2018). A bayesian framework to optimize double band spectra spatial filters for motor imagery classification. In *2018 ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 871–875).

Subasi, A., & Gursoy, M. I. (2010). Eeg signal classification using pca, ica, lda and support vector machines. *Expert systems with applications*, *37*(12), 8659–8666.

Suk, H.-I., & Lee, S.-W. (2013). A novel bayesian framework for discriminative feature extraction in brain-computer interfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(2), 286–299.

Sun, S., & Zhang, C. (2006). Adaptive feature extraction for eeg signal classification. *Medical and Biological Engineering and Computing*, *44*(10), 931–935.

Townsend, G., Graimann, B., & Pfurtscheller, G. (2006). A comparison of common spatial patterns with complex band power features in a four-class bci experiment. *IEEE Transactions on Biomedical Engineering*, *53*(4), 642–651.

Van Erp, J., Lotte, F., & Tangermann, M. (2012). Brain-computer interfaces: beyond medical applications. *Computer*, *45*(4), 26–34.

Vaughan, T. M., Heetderks, W., Trejo, L., Rymer, W., Weinrich, M., Moore, M., ... others (2003). *Brain-computer interface technology: a review of the second international meeting.*

Vidal, J. J. (1973). Toward direct brain-computer communication. *Annual review of Biophysics and Bioengineering*, *2*(1), 157–180.

Wolpaw, J., & Wolpaw, E. W. (2012). *Brain-computer interfaces: principles and practice*. OUP USA.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical neurophysiology*, *113*(6), 767–791.

Wong, T.-T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, *48*(9), 2839–2846.

Xie, X., Yu, Z. L., Lu, H., Gu, Z., & Li, Y. (2017). Motor imagery classification based on bilinear sub-manifold learning of symmetric positive-definite matrices. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(6), 504–516.

Yang, B., He, L., Wang, Q., Song, C., & Zhang, Y. (2014). A preprocessing method of eeg based on emd-ica in bci. In *International conference on life system modeling and simulation and international conference on intelligent computing for sustainable energy and environment* (pp. 1–12).

Yu, X., Chum, P., & Sim, K.-B. (2014). Analysis the effect of pca for feature reduction in non-stationary eeg based motor imagery of bci system. *Optik-International Journal for Light and Electron Optics*, *125*(3), 1498–1502.

Yuan, H., & He, B. (2014). Brain–computer interfaces using sensorimotor rhythms: current state and future perspectives. *IEEE Transactions on Biomedical Engineering*, *61*(5), 1425–1435.

Zander, T. O., & Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of neural engineering*, *8*(2), 025005.